Wiley Classics Library

COURANT & HILBERT

Methods of Mathematical Physics Volume I

METHODS OF MATHEMATICAL PHYSICS

VOLUME I

METHODS OF MATHEMATICAL PHYSICS

By R. COURANT and D. HILBERT

First English edition
Translated and Revised from the German Original

VOLUME I

INTERSCIENCE PUBLISHERS, INC., NEW YORK
a division of John Wiley & Sons, New York · London · Sydney

Copyright 1937 by Julius Springer, Berlin
vested in the Attorney General
of the United States, pursuant to law.
Manufactured and sold under License A-1556
Copyright 1953 by Interscience Publishers, Inc.

Library of Congress Catalog Card Number 53-7164

517.3
C04.C3

SEVENTH PRINTING, DECEMBER, 1966

PRINTED IN THE UNITED STATES OF AMERICA

142615

PREFACE

The first German edition of this volume was published by Julius Springer, Berlin, in 1924. A second edition, revised and improved with the help of K. O. Friedrichs, R. Luneburg, F. Rellich, and other unselfish friends, followed in 1930. The second volume appeared in 1938. In the meantime I had been forced to leave Germany and was fortunate and grateful to be given the opportunities open in the United States. During the Second World War the German book became unavailable and later was even suppressed by the National Socialist rulers of Germany. Thus the survival of the book was secured when the United States Government seized the copyright and licensed a reprint issued by Interscience Publishers, New York. Such a license also had to be obtained from the Alien Property Custodian for the present English edition.

This edition follows the German original fairly closely but contains a large number of additions and modifications. I have had to postpone a plan to completely rewrite and modernize the book in collaboration with K. O. Friedrichs, because the pressure for publication of an English "Courant-Hilbert" has become irresistible. Even so, it is hoped that the work in its present form will be useful to mathematicians and physicists alike, as the numerous demands from all sides seem to indicate.

The objective of the book can still today be expressed almost as in the preface to the first German edition. "Since the seventeenth century, physical intuition has served as a vital source for mathematical problems and methods. Recent trends and fashions have, however, weakened the connection between mathematics and physics; mathematicians, turning away from the roots of mathematics in intuition, have concentrated on refinement and emphasized the postulational side of mathematics, and at times have overlooked the unity of their science with physics and other fields. In many cases, physicists have ceased to appreciate the attitudes of mathematicians. This rift is unquestionably a serious threat to science as a whole; the broad stream of scientific development may split into smaller and

smaller rivulets and dry out. It seems therefore important to direct our efforts toward reuniting divergent trends by clarifying the common features and interconnections of many distinct and diverse scientific facts. Only thus can the student attain some mastery of the material and the basis be prepared for further organic development of research.

"The present work is designed to serve this purpose for the field of mathematical physics. Mathematical methods originating in problems of physics are developed and the attempt is made to shape results into unified mathematical theories. Completeness is not attempted, but it is hoped that access to a rich and important field will be facilitated by the book.

"The responsibility for the present book rests with me. Yet the name of my teacher, colleague, and friend, D. Hilbert, on the title page seems justified by the fact that much material from Hilbert's papers and lectures has been used, as well as by the hope that the book expresses some of Hilbert's spirit, which has had such a decisive influence on mathematical research and education."

I am greatly indebted to many helpers in all phases of the task of preparing this edition: to Peter Ceike, Ernest Courant, and Anneli Lax, who provided most of the first draft of the translation; to Hanan Rubin and Herbert Kranzer, who have given constructive criticism; to Wilhelm Magnus, who is responsible for the appendix to Chapter VII; and to Natascha Artin and Lucile Gardner, who carried the burden of the editorial work. Most cordial thanks also are due to Interscience Publishers for their patient and helpful attitude and to my old friend and publisher, Dr. Ferdinand Springer in Heidelberg, the great pioneer of modern scientific publishing, for his sympathetic understanding of the situation, which has so greatly changed since the old days of our close cooperation.

R. COURANT

New Rochelle, New York
June 1953

CONTENTS

I. The Algebra of Linear Transformations and Quadratic Forms 1

§1. Linear equations and linear transformations 1
1. Vectors 1
2. Orthogonal systems of vectors. Completeness 3
3. Linear transformations. Matrices 5
4. Bilinear, quadratic, and Hermitian forms 11
5. Orthogonal and unitary transformations 14
§2. Linear transformations with a linear parameter 17
§3. Transformation to principal axes of quadratic and Hermitian forms 23
1. Transformation to principal axes on the basis of a maximum principle 23
2. Eigenvalues 26
3. Generalization to Hermitian forms 28
4. Inertial theorem for quadratic forms 28
5. Representation of the resolvent of a form 29
6. Solution of systems of linear equations associated with forms 30
§4. Minimum-maximum property of eigenvalues 31
1. Characterization of eigenvalues by a minimum-maximum problem 31
2. Applications. Constraints 33
§5. Supplement and problems 34
1. Linear independence and the Gram determinant 34
2. Hadamard's inequality for determinants 36
3. Generalized treatment of canonical transformations 37
4. Bilinear and quadratic forms of infinitely many variables 41
5. Infinitesimal linear transformations 41
6. Perturbations 42
7. Constraints 44
8. Elementary divisors of a matrix or a bilinear form 45
9. Spectrum of a unitary matrix 46
References 47

II. Series Expansions of Arbitrary Functions 48

§1. Orthogonal systems of functions 49
1. Definitions 49
2. Orthogonalization of functions 50
3. Bessel's inequality. Completeness relation. Approximation in the mean 51

4. Spectral decomposition by Fourier series and integrals ... 99
5. Dense systems of functions ... 100
6. A Theorem of H. Müntz on the completeness of powers ... 102
7. Fejér's summation theorem ... 102
8. The Mellin inversion formulas ... 103
9. The Gibbs phenomenon ... 105
10. A theorem on Gram's determinant ... 107
11. Application of the Lebesgue integral ... 108
References ... 111

III. Linear Integral Equations ... 112

§1. Introduction ... 112
1. Notation and basic concepts ... 112
2. Functions in integral representation ... 113
3. Degenerate kernels ... 114
§2. Fredholm's theorems for degenerate kernels ... 115
§3. Fredholm's theorems for arbitrary kernels ... 118
§4. Symmetric kernels and their eigenvalues ... 122
1. Existence of an eigenvalue of a symmetric kernel ... 122
2. The totality of eigenfunctions and eigenvalues ... 126
3. Maximum-minimum property of eigenvalues ... 132
§5. The expansion theorem and its applications ... 134
1. Expansion theorem ... 134
2. Solution of the inhomogeneous linear integral equation ... 136
3. Bilinear formula for iterated kernels ... 137
4. Mercer's theorem ... 138
§6. Neumann series and the reciprocal kernel ... 140
§7. The Fredholm formulas ... 142
§8. Another derivation of the theory ... 147
1. A lemma ... 147
2. Eigenfunctions of a symmetric kernel ... 148
3. Unsymmetric kernels ... 150
4. Continuous dependence of eigenvalues and eigenfunctions on the kernel ... 151
§9. Extensions of the theory ... 152
§10. Supplement and problems for Chapter III ... 153
1. Problems ... 153
2. Singular integral equations ... 154
3. E. Schmidt's derivation of the Fredholm theorems ... 155
4. Enskog's method for solving symmetric integral equations ... 156
5. Kellogg's method for the determination of eigenfunctions ... 156
6. Symbolic functions of a kernel and their eigenvalues ... 157
7. Example of an unsymmetric kernel without null solutions ... 157
8. Volterra integral equation ... 158
9. Abel's integral equation ... 158
10. Adjoint orthogonal systems belonging to an unsymmetric kernel ... 159
11. Integral equations of the first kind ... 159

12. Method of infinitely many variables... 160
13. Minimum properties of eigenfunctions 161
14. Polar integral equations..... 161
15. Symmetrizable kernels..... 161
16. Determination of the resolvent kernel by functional equations..... 162
17. Continuity of definite kernels... 162
18. Hammerstein's theorem. 162
References 162

IV. The Calculus of Variations 164

§1. Problems of the calculus of variations. 164
1. Maxima and minima of functions. 164
2. Functionals..... 167
3. Typical problems of the calculus of variations..... 169
4. Characteristic difficulties of the calculus of variations..... 173
§2. Direct solutions..... 174
1. The isoperimetric problem..... 174
2. The Rayleigh-Ritz method. Minimizing sequences..... 175
3. Other direct methods. Method of finite differences. Infinitely many variables..... 176
4. General remarks on direct methods of the calculus of variations..... 182
§3. The Euler equations..... 183
1. "Simplest problem" of the variational calculus..... 184
2. Several unknown functions. 187
3. Higher derivatives..... 190
4. Several independent variables..... 191
5. Identical vanishing of the Euler differential expression.. 193
6. Euler equations in homogeneous form..... 196
7. Relaxing of conditions. Theorems of du Bois-Reymond and Haar..... 199
8. Variational problems and functional equations.... 205
§4. Integration of the Euler differential equation. 206
§5. Boundary conditions..... 208
1. Natural boundary conditions for free boundaries... 208
2. Geometrical problems. Transversality..... 211
§6. The second variation and the Legendre condition.. 214
§7. Variational problems with subsidiary conditions. 216
1. Isoperimetric problems.. 216
2. Finite subsidiary conditions..... 219
3. Differential equations as subsidiary conditions.. 221
§8. Invariant character of the Euler equations..... 222
1. The Euler expression as a gradient in function space. Invariance of the Euler expression..... 222
2. Transformation of Δu. Spherical coordinates.. 224
3. Ellipsoidal coordinates..... 226

§9. Transformation of variational problems to canonical and involutory form ... 231
1. Transformation of an ordinary minimum problem with subsidiary conditions ... 231
2. Involutory transformation of the simplest variational problems ... 233
3. Transformation of variational problems to canonical form ... 238
4. Generalizations ... 240
§10. Variational calculus and the differential equations of mathematical physics ... 242
1. General remarks ... 242
2. The vibrating string and the vibrating rod ... 244
3. Membrane and plate ... 246
§11. Reciprocal quadratic variational problems ... 252
§12. Supplementary remarks and exercises ... 257
1. Variational problem for a given differential equation ... 257
2. Reciprocity for isoperimetric problems ... 258
3. Circular light rays ... 258
4. The problem of Dido ... 258
5. Examples of problems in space ... 258
6. The indicatrix and applications ... 258
7. Variable domains ... 260
8. E. Noether's theorem on invariant variational problems. Integrals in particle mechanics ... 262
9. Transversality for multiple integrals ... 266
10. Euler's differential expressions on surfaces ... 267
11. Thomson's principle in electrostatics ... 267
12. Equilibrium problems for elastic bodies. Castigliano's principle ... 268
13. The variational problem of buckling ... 272
References ... 274

V. **Vibration and Eigenvalue Problems** ... 275

§1. Preliminary remarks about linear differential equations ... 275
1. Principle of superposition ... 275
2. Homogeneous and nonhomogeneous problems. Boundary conditions ... 277
3. Formal relations. Adjoint differential expressions. Green's formulas ... 277
4. Linear functional equations as limiting cases and analogues of systems of linear equations ... 280
§2. Systems of a finite number of degrees of freedom ... 281
1. Normal modes of vibration. Normal coordinates. General theory of motion ... 282
2. General properties of vibrating systems ... 285
§3. The vibrating string ... 286
1. Free motion of the homogeneous string ... 287

2. Forced motion ... 289
3. The general nonhomogeneous string and the Sturm-Liouville eigenvalue problem ... 291
§4. The vibrating rod ... 295
§5. The vibrating membrane ... 297
1. General eigenvalue problem for the homogeneous membrane 297
2. Forced motion ... 300
3. Nodal lines ... 300
4. Rectangular membrane ... 300
5. Circular membrane. Bessel functions ... 302
6. Nonhomogeneous membrane ... 306
§6. The vibrating plate ... 307
1. General remarks ... 307
2. Circular boundary ... 307
§7. General remarks on the eigenfunction method ... 308
1. Vibration and equilibrium problems ... 308
2. Heat conduction and eigenvalue problems ... 311
§8. Vibration of three-dimensional continua. Separation of variables ... 313
§9. Eigenfunctions and the boundary value problem of potential theory ... 315
1. Circle, sphere, and spherical shell ... 315
2. Cylindrical domain ... 319
3. The Lamé problem ... 319
§10. Problems of the Sturm-Liouville type. Singular boundary points ... 324
1. Bessel functions ... 324
2. Legendre functions of arbitrary order ... 325
3. Jacobi and Tchebycheff polynomials ... 327
4. Hermite and Laguerre polynomials ... 328
§11. The asymptotic behavior of the solutions of Sturm-Liouville equations ... 331
1. Boundedness of the solution as the independent variable tends to infinity ... 331
2. A sharper result. (Bessel functions) ... 332
3. Boundedness as the parameter increases ... 334
4. Asymptotic representation of the solutions ... 335
5. Asymptotic representation of Sturm-Liouville eigenfunctions ... 336
§12. Eigenvalue problems with a continuous spectrum ... 339
1. Trigonometric functions ... 340
2. Bessel functions ... 340
3. Eigenvalue problem of the membrane equation for the infinite plane ... 341
4. The Schrödinger eigenvalue problem ... 341
§13. Perturbation theory ... 343

1. Simple eigenvalues ... 344
2. Multiple eigenvalues ... 346
3. An example ... 348
§14. Green's function (influence function) and reduction of differential equations to integral equations ... 351
1. Green's function and boundary value problem for ordinary differential equations ... 351
2. Construction of Green's function; Green's function in the generalized sense ... 354
3. Equivalence of integral and differential equations ... 358
4. Ordinary differential equations of higher order ... 362
5. Partial differential equations ... 363
§15. Examples of Green's function ... 371
1. Ordinary differential equations ... 371
2. Green's function for Δu: circle and sphere ... 377
3. Green's function and conformal mapping ... 377
4. Green's function for the potential equation on the surface of a sphere ... 378
5. Green's function for $\Delta u = 0$ in a rectangular parallelepiped ... 378
6. Green's function for Δu in the interior of a rectangle ... 384
7. Green's function for a circular ring ... 386
§16. Supplement to Chapter V ... 388
1. Examples for the vibrating string ... 388
2. Vibrations of a freely suspended rope; Bessel functions ... 390
3. Examples for the explicit solution of the vibration equation. Mathieu functions ... 391
4. Boundary conditions with parameters ... 392
5. Green's tensors for systems of differential equations ... 393
6. Analytic continuation of the solutions of the equation $\Delta u + \lambda u = 0$ ... 395
7. A theorem on the nodal curves of the solutions of $\Delta u + \lambda u = 0$ ... 395
8. An example of eigenvalues of infinite multiplicity ... 395
9. Limits for the validity of the expansion theorems ... 395
References ... 396

VI. Application of the Calculus of Variations to Eigenvalue Problems ... 397

§1. Extremum properties of eigenvalues ... 398
1. Classical extremum properties ... 398
2. Generalizations ... 402
3. Eigenvalue problems for regions with separate components ... 405
4. The maximum-minimum property of eigenvalues ... 405
§2. General consequences of the extremum properties of the eigenvalues ... 407
1. General theorems ... 407

2. Infinite growth of the eigenvalues ... 412
3. Asymptotic behavior of the eigenvalues in the Sturm-Liouville problem ... 414
4. Singular differential equations ... 415
5. Further remarks concerning the growth of eigenvalues. Occurrence of negative eigenvalues ... 416
6. Continuity of eigenvalues ... 418
§3. Completeness and expansion theorems ... 424
1. Completeness of the eigenfunctions ... 424
2. The expansion theorem ... 426
3. Generalization of the expansion theorem ... 427
§4. Asymptotic distribution of eigenvalues ... 429
1. The equation $\Delta u + \lambda u = 0$ for a rectangle ... 429
2. The equation $\Delta u + \lambda u = 0$ for domains consisting of a finite number of squares or cubes ... 431
3. Extension to the general differential equation $L[u] + \lambda \rho u = 0$ ... 434
4. Asymptotic distribution of eigenvalues for an arbitrary domain ... 436
5. Sharper form of the laws of asymptotic distribution of eigenvalues for the differential equation $\Delta u + \lambda u = 0$ ... 443
§5. Eigenvalue problems of the Schrödinger type ... 445
§6. Nodes of eigenfunctions ... 451
§7. Supplementary remarks and problems ... 455
1. Minimizing properties of eigenvalues. Derivation from completeness ... 455
2. Characterization of the first eigenfunction by absence of nodes ... 458
3. Further minimizing properties of eigenvalues ... 459
4. Asymptotic distribution of eigenvalues ... 460
5. Parameter eigenvalue problems ... 460
6. Boundary conditions containing parameters ... 461
7. Eigenvalue problems for closed surfaces ... 461
8. Estimates of eigenvalues when singular points occur ... 461
9. Minimum theorems for the membrane and plate ... 463
10. Minimum problems for variable mass distribution ... 463
11. Nodal points for the Sturm-Liouville problem. Maximum-minimum principle ... 463
References ... 464

VII. Special Functions Defined by Eigenvalue Problems ... 466

§1. Preliminary discussion of linear second order differential equations ... 466
§2. Bessel functions ... 467
1. Application of the integral transformation ... 468
2. Hankel functions ... 469
3. Bessel and Neumann functions ... 471
4. Integral representations of Bessel functions ... 474

5. Another integral representation of the Hankel and Bessel functions ... 476
6. Power series expansion of Bessel functions ... 482
7. Relations between Bessel functions ... 485
8. Zeros of Bessel functions ... 492
9. Neumann functions ... 496
§3. Legendre functions ... 501
1. Schläfli's integral ... 501
2. Integral representations of Laplace ... 503
3. Legendre functions of the second kind ... 504
4. Associated Legendre functions. (Legendre functions of higher order.) ... 505
§4. Application of the method of integral transformation to Legendre, Tchebycheff, Hermite, and Laguerre equations ... 506
1. Legendre functions ... 506
2. Tchebycheff functions ... 507
3. Hermite functions ... 508
4. Laguerre functions ... 509
§5. Laplace spherical harmonics ... 510
1. Determination of $2n + 1$ spherical harmonics of n-th order ... 511
2. Completeness of the system of functions. 512
3. Expansion theorem ... 513
4. The Poisson integral ... 513
5. The Maxwell-Sylvester representation of spherical harmonics ... 514
§6. Asymptotic expansions ... 522
1. Stirling's formula ... 522
2. Asymptotic calculation of Hankel and Bessel functions for large values of the arguments ... 524
3. The saddle point method ... 526
4. Application of the saddle point method to the calculation of Hankel and Bessel functions for large parameter and large argument ... 527
5. General remarks on the saddle point method. 532
6. The Darboux method ... 532
7. Application of the Darboux method to the asymptotic expansion of Legendre polynomials ... 533
§7. Appendix to Chapter VII. Transformation of Spherical Harmonics ... 535
1. Introduction and notation ... 535
2. Orthogonal transformations ... 536
3. A generating function for spherical harmonics. 539
4. Transformation formula ... 542
5. Expressions in terms of angular coordinates ... 543

Additional Bibliography ... 547

Index ... 551

CHAPTER I

The Algebra of Linear Transformations and Quadratic Forms

In the present volume we shall be concerned with many topics in mathematical analysis which are intimately related to the theory of linear transformations and quadratic forms. A brief résumé of pertinent aspects of this field will, therefore, be given in Chapter I. The reader is assumed to be familiar with the subject in general.

§1. *Linear Equations and Linear Transformations*

1. Vectors. The results of the theory of linear equations can be expressed concisely by the notation of vector analysis. A system of n real numbers $x_1, x_2, \cdots, x_n$ is called an *n-dimensional vector* or a vector in n-dimensional space and denoted by the bold face letter $\mathbf{x}$; the numbers x_i $(i = 1, \cdots, n)$ are called the *components* of the vector $\mathbf{x}$. If all components vanish, the vector is said to be zero or the *null vector*; for $n = 2$ or $n = 3$ a vector can be interpreted geometrically as a "position vector" leading from the origin to the point with the rectangular coordinates x_i. For $n > 3$ geometrical visualization is no longer possible but geometrical terminology remains suitable.

Given two arbitrary real numbers λ and μ, the vector $\lambda\mathbf{x} + \mu\mathbf{y} = \mathbf{z}$ is defined as the vector whose components z_i are given by $z_i = \lambda x_i + \mu y_i$. Thus, in particular, the sum and difference of two vectors are defined.

The number

$$(1) \qquad \mathbf{x}\cdot\mathbf{y} = x_1y_1 + \cdots + x_ny_n = y_1x_1 + \cdots + y_nx_n = \mathbf{y}\cdot\mathbf{x}$$

is called the "*inner product*" of the vectors $\mathbf{x}$ and $\mathbf{y}$.

Occasionally we shall call the inner product $\mathbf{x}\cdot\mathbf{y}$ the *component of the vector* $\mathbf{y}$ *with respect to* $\mathbf{x}$ or vice versa.

If the inner product $\mathbf{x}\cdot\mathbf{y}$ vanishes we say that the vectors $\mathbf{x}$ and $\mathbf{y}$ are *orthogonal*; for $n = 2$ and $n = 3$ this terminology has an imme-

diate geometrical meaning. The inner product $\mathbf{x}\cdot\mathbf{x} = \mathbf{x}^2$ of a vector with itself plays a special role; it is called the *norm* of the vector. The positive square root of $\mathbf{x}^2$ is called the *length* of the vector and denoted by $|\mathbf{x}| = \sqrt{\mathbf{x}^2}$. A vector whose length is unity is called a *normalized vector* or *unit vector*.

The following inequality is satisfied by the inner product of two vectors $\mathbf{a} = (a_1, \cdots, a_n)$ and $\mathbf{b} = (b_1, \cdots, b_n)$:

$$(\mathbf{a}\cdot\mathbf{b})^2 \leq \mathbf{a}^2\mathbf{b}^2$$

or, without using vector notation,

$$\left(\sum_{i=1}^{n} a_i b_i\right)^2 \leq \left(\sum_{i=1}^{n} a_i^2\right)\left(\sum_{i=1}^{n} b_i^2\right),$$

where the equality holds if and only if the a_i and the b_i are proportional, i.e. if a relation of the form $\lambda\mathbf{a} + \mu\mathbf{b} = 0$ with $\lambda^2 + \mu^2 \neq 0$ is satisfied.

The proof of this "*Schwarz inequality*"[1] follows from the fact that the roots of the quadratic equation

$$\sum_{i=1}^{n} (a_i x + b_i)^2 = x^2 \sum_{i=1}^{n} a_i^2 + 2x \sum_{i=1}^{n} a_i b_i + \sum_{i=1}^{n} b_i^2 = 0$$

for the unknown x can never be real and distinct, but must be imaginary, unless the a_i and b_i are proportional. The Schwarz inequality is merely an expression of this fact in terms of the discriminant of the equation. Another proof of the Schwarz inequality follows immediately from the identity

$$\sum_{i=1}^{n} a_i^2 \sum_{i=1}^{n} b_i^2 - \left(\sum_{i=1}^{n} a_i b_i\right)^2 = \tfrac{1}{2} \sum_{j=1}^{n} \sum_{k=1}^{n} (a_j b_k - a_k b_j)^2.$$

Vectors $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_m$ are said to be *linearly dependent* if a set of numbers $\lambda_1, \lambda_2, \cdots, \lambda_m$ (not all equal to zero) exists such that the vector equation

$$\lambda_1\mathbf{x}_1 + \cdots + \lambda_m\mathbf{x}_m = 0$$

is satisfied, i.e. such that all the components of the vector on the left vanish. Otherwise the vectors are said to be *linearly independent*.

The n vectors $\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_n$ in n-dimensional space whose com-

[1] This relation was, as a matter of fact, used by Cauchy before Schwarz.

ponents are given, respectively, by the first, second, $\cdots$, and n-th rows of the array

$$\begin{matrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 1, \end{matrix}$$

form a system of n linearly independent vectors. For, if a relation $\lambda_1\mathbf{e}_1 + \cdots + \lambda_n\mathbf{e}_n = 0$ were satisfied, we could multiply[1] this relation by $\mathbf{e}_h$ and obtain $\lambda_h = 0$ for every h, since $\mathbf{e}_h^2 = 1$ and $\mathbf{e}_h \cdot \mathbf{e}_k = 0$ if $h \neq k$. Thus, systems of n linearly independent vectors certainly exist. However, for any $n + 1$ vectors $\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_{n+1}$ (in n-dimensional space) there is at least one linear equation of the form

$$\mu_1\mathbf{u}_1 + \cdots + \mu_{n+1}\mathbf{u}_{n+1} = 0,$$

with coefficients that do not all vanish, since n homogeneous linear equations

$$\sum_{i=1}^{n+1} u_{ik}\mu_i = 0 \qquad (k = 1, \cdots, n)$$

for the $n + 1$ unknowns $\mu_1, \mu_2, \cdots, \mu_{n+1}$ always have at least one nontrivial solution (cf. subsection 3).

2. Orthogonal Systems of Vectors. Completeness. The above "coordinate vectors" $\mathbf{e}_i$ form a particular system of *orthogonal unit vectors*. In general a system of n orthogonal unit vectors $\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_n$ is defined as a system of vectors of unit length satisfying the relations

$$\mathbf{e}_h^2 = 1, \qquad \mathbf{e}_h \cdot \mathbf{e}_k = 0 \qquad (h \neq k)$$

for $h, k, = 1, 2, \cdots, n$. As above, we see that the n vectors $\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_n$ are linearly independent.

If $\mathbf{x}$ is an arbitrary vector, a relation of the form

$$c_0\mathbf{x} - c_1\mathbf{e}_1 - \cdots - c_n\mathbf{e}_n = 0$$

with constants c_i that do not all vanish must hold; for, as we have seen, any $n + 1$ vectors are linearly dependent. Since the $\mathbf{e}_i$ are linearly independent, c_0 cannot be zero; we may therefore, without

[1] To multiply two vectors is to take their inner product.

loss of generality, take it to be equal to unity. Every vector $\mathbf{x}$ can thus be expressed in terms of a system of orthogonal unit vectors in the form

$$\mathbf{x} = c_1\mathbf{e}_1 + \cdots + c_n\mathbf{e}_n. \tag{2}$$

The coefficients c_i, the *components of* $\mathbf{x}$ *with respect to the system* $\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_n$, may be found by multiplying (2) by each of the vectors $\mathbf{e}_i$; they are

$$c_i = \mathbf{x}\cdot\mathbf{e}_i.$$

From any arbitrary system of m linearly independent vectors $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_m$, we may, by the following *orthogonalization process* due to E. Schmidt, obtain a system of m orthogonal unit vectors $\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_m$: First set $\mathbf{e}_1 = \mathbf{v}_1/|\mathbf{v}_1|$. Then choose a number c_1' in such a way that $\mathbf{v}_2 - c_1'\mathbf{e}_1$ is orthogonal to $\mathbf{e}_1$, i.e. set $c_1' = \mathbf{v}_2\cdot\mathbf{e}_1$. Since $\mathbf{v}_1$ and $\mathbf{v}_2$, and therefore $\mathbf{e}_1$ and $\mathbf{v}_2$, are linearly independent, the vector $\mathbf{v}_2 - c_1'\mathbf{e}_1$ is different from zero. We may then divide this vector by its length obtaining a unit vector $\mathbf{e}_2$ which is orthogonal to $\mathbf{e}_1$. We next find two numbers c_1'', c_2'' such that $\mathbf{v}_3 - c_1''\mathbf{e}_1 - c_2''\mathbf{e}_2$ is orthogonal to both $\mathbf{e}_1$ and $\mathbf{e}_2$, i.e. we set $c_1'' = \mathbf{v}_3\cdot\mathbf{e}_1$ and $c_2'' = \mathbf{v}_3\cdot\mathbf{e}_2$. This vector is again different from zero and can, therefore, be normalized; we divide it by its length and obtain the unit vector $\mathbf{e}_3$. By continuing this procedure we obtain the desired orthogonal system.

For $m < n$ the resulting orthogonal system is called *incomplete*, and if $m = n$ we speak of a *complete orthogonal system*. Let us denote the components of a vector $\mathbf{x}$ with respect to $\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_m$ by $c_1, c_2, \cdots, c_m$ as before. The self-evident inequality

$$(\mathbf{x} - c_1\mathbf{e}_1 - \cdots - c_m\mathbf{e}_m)^2 \geq 0$$

is satisfied. Evaluating the left side term by term according to the usual rules of algebra (which hold for vectors if the inner product of two vectors is used whenever two vectors are multiplied), we find

$$\mathbf{x}^2 - 2\mathbf{x}\cdot\sum_{i=1}^{m} c_i\mathbf{e}_i + \sum_{i=1}^{m} c_i^2 = \mathbf{x}^2 - 2\sum_{i=1}^{m} c_i^2 + \sum_{i=1}^{m} c_i^2 \geq 0$$

or

$$\mathbf{x}^2 \geq \sum_{i=1}^{m} c_i^2, \tag{3}$$

where $m \leq n$ and $c_i = \mathbf{x} \cdot \mathbf{e}_i$; the following equality holds for $m = n$:

$$\mathbf{x}^2 = \sum_{i=1}^{m} c_i^2. \tag{4}$$

Relations (3) and (4)—(4) expresses the theorem of Pythagoras in vector notation—have an intuitive significance for $n \leq 3$; they are called, respectively, *Bessel's inequality* and the *completeness relation*. Relation (4), if it is satisfied for every vector $\mathbf{x}$, does in fact indicate that the given orthogonal system is complete since (4) could not be satisfied for a unit vector orthogonal to all vectors $\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_m$, and such a vector necessarily exists if $m < n$.

The completeness relation may also be expressed in the more general form

$$\mathbf{x} \cdot \mathbf{x}' = \sum_{i=1}^{m} c_i c_i', \tag{5}$$

which follows from the orthogonality of the $\mathbf{e}_i$.

So far these algebraic relations are all purely formal. Their significance lies in the fact that they occur again in a similar manner in transcendental problems of analysis.

3. Linear Transformations. Matrices. A system of n linear equations

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= y_1, \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= y_2, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots & \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n &= y_n, \end{aligned} \tag{6}$$

with coefficients a_{ik}, assigns a unique set of quantities $y_1, y_2, \cdots, y_n$ to every set of quantities $x_1, x_2, \cdots, x_n$. Such an assignment is called a *linear transformation* of the set $x_1, x_2, \cdots, x_n$ into the set $y_1, y_2, \cdots, y_n$, or, briefly, of the vector $\mathbf{x}$ into the vector $\mathbf{y}$. This transformation is clearly linear since the vector $\lambda_1 y_1 + \lambda_2 y_2$ corresponds to the vector $\lambda_1 \mathbf{x}_1 + \lambda_2 \mathbf{x}_2$.

The most important problem in connection with linear transformations is the p· oblem of inversion, the question, in other words, of the

existence of a solution of a system of linear equations. The answer is given by the following fundamental theorem of the theory of linear equations, whose proof we assume to be known:
For the system of equations

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= y_1, \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= y_2, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots & \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n &= y_n, \end{aligned}$$

or, briefly,

$$\sum_{k=1}^{n} a_{ik}x_k = y_i \qquad (i = 1, \cdots, n), \tag{7}$$

with given coefficients a_{ik}, *the following alternative holds: Either it has one and only one solution* $\mathfrak{x}$ *for each arbitrarily given vector* $\mathfrak{y}$, *in particular the solution* $\mathfrak{x} = 0$ *for* $\mathfrak{y} = 0$; *or, alternatively, the homogeneous equations arising from* (7) *for* $\mathfrak{y} = 0$ *have a positive number* ρ *of nontrivial (not identically zero) linearly independent solutions* $\mathfrak{x}_1, \mathfrak{x}_2, \cdots, \mathfrak{x}_\rho$, *which may be assumed to be normalized. In the latter case the "transposed" homogeneous system of equations*

$$\sum_{k=1}^{n} a'_{ik}x'_k = 0 \qquad (i = 1, \cdots, n), \tag{8}$$

where $a'_{ik} = a_{ki}$, *also has exactly* ρ *linearly independent nontrivial solutions* $\mathfrak{x}'_1, \mathfrak{x}'_2, \cdots, \mathfrak{x}'_\rho$. *The inhomogeneous system* (7) *then possesses solutions for just those vectors* $\mathfrak{y}$ *which are orthogonal to* $\mathfrak{x}'_1, \mathfrak{x}'_2, \cdots, \mathfrak{x}'_\rho$. *These solutions are determined only to within an additive term which is an arbitrary solution of the homogeneous system of equations, i.e. if* $\mathfrak{x}$ *is a solution of the inhomogeneous system and* $\mathfrak{x}_\sigma$ *is any solution of the homogeneous system, then* $\mathfrak{x} + \mathfrak{x}_\sigma$ *is also a solution of the inhomogeneous system.*

In this formulation of the fundamental theorem reference to the theory of determinants has been avoided. Later, to obtain explicit expressions for the solutions of the system of equations, determinants will be required.

The essential features of such a linear transformation are contained in the array of coefficients or *matrix* of the equations (7):

$$A = (a_{ik}) = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} \tag{9}$$

with the determinant

$$\mathsf{A} = |a_{ik}| = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix}.$$

It is sometimes useful to denote the transformation itself (also called *tensor*[1] or *operator*) by a special letter **A**. The elements a_{ik} of the matrix A are called the components of the tensor. The linear transformation (7) may be regarded as a "*multiplication*" of the tensor **A** by the vector **x**, written symbolically in the form

$$\mathbf{Ax} = \mathbf{y}.$$

Many results in the algebra of linear transformations may be expressed concisely in terms of matrices or tensors, once certain simple rules and definitions known as *matrix algebra* have been introduced. First we define matrix multiplication; this concept arises if we suppose that the vector **x**, which is transformed in equations (7), is itself the product of a tensor **B** with components b_{ik} and another vector **w**:

$$\sum_{j=1}^{n} b_{kj} w_j = x_k \qquad (k = 1, \cdots, n).$$

Multiplying **w** by a tensor **C** we obtain the vector **y**. The matrix C which corresponds to the tensor **C** is obtained from A and B by the rule of *matrix multiplication*, $C = AB$, which states that the element c_{ij} is the inner product of the i-th row of A and the j-th column of B:

$$c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj} \qquad (i, j = 1, \cdots, n). \tag{10}$$

[1] In modern usage the term "operator" is customary to denote linear transformations.

The tensor or transformation **C** is therefore called the inner product or simply the *product of the tensors or transformations* **A** and **B**. Henceforth tensors and the equivalent matrices will not be distinguished from each other. Note that matrix products obey the *associative law*

$$(AB)C = A(BC),$$

so that the product $A_1A_2 \cdots A_h$ of any number of matrices written in a fixed order has a unique meaning. For $A_1 = A_2 = \cdots = A_h = A$ we write this product as the h-th power A^h of the matrix A. It is, on the other hand, essential to note that the *commutative law* of multiplication is in general *not* valid; AB, in other words, differs in general from BA. Finally the matrix $\lambda A + \mu B$ is defined as the matrix whose elements are $\lambda a_{ik} + \mu b_{ik}$; thus the null matrix is the matrix in which all components vanish.[1] The validity of the *distributive law*

$$(A + B)C = AC + BC$$

is immediately evident.

The *unit matrix* is defined by

$$E = (e_{ik}) = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}.$$

It is characterized by the fact that the equation

$$AE = EA = A$$

holds for an arbitrary matrix A. The unit matrix corresponds to the *identity transformation*

$$x_i = y_i \qquad (i = 1, \cdots, n).$$

The zero-th power of every matrix A is defined as the unit matrix:

$$A^0 = E.$$

[1] Note that in matrix algebra it does not necessarily follow from the matrix equation $AB = (0)$ that one of the two factors vanishes, as can be seen from the example $A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, B = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$.

Since the powers A^h of a matrix are defined, we can also define polynomials whose argument is a matrix. Thus, if

$$f(x) = a_0 + a_1 x + \cdots + a_m x^m$$

is a polynomial of the m-th degree in the variable x, then $f(A)$ is defined by

$$f(A) = a_0 E + a_1 A + \cdots + a_m A^m$$

as a (symbolic) polynomial in the matrix A. This definition of a matrix as a function $f(A)$ of A can even, on occasion, be extended to functions which are not polynomials but which can be expressed as power series. The matrix e^A, for example, may be defined by

$$B = e^A = E + A + \frac{A^2}{2!} + \frac{A^3}{3!} + \cdots = \sum_{\nu=0}^{\infty} \frac{A^\nu}{\nu!}.$$

Note that in such a series one first considers the sum of the first N terms and then investigates whether each of the n^2 elements of the resulting matrix converges to a limit with increasing N; if this is the case, the matrix formed from the n^2 limiting values is considered to be the sum of the series. In the particular case of the matrix e^A it turns out, as will be shown below, that the series always converges.

A particularly important relation is obtained for a matrix S defined by a *geometric series* with partial sums S_m given by

$$S_m = E + A + A^2 + \cdots + A^m.$$

Multiplying the equation which defines S_m by A, we obtain the equation

$$S_m A + E = S_m + A^{m+1},$$

from which it follows that

$$S_m(E - A) = E - A^{m+1}.$$

Now if the matrix S_m approaches a limit S with increasing m, so that A^{m+1} tends to zero, we obtain the relation

$$S(E - A) = E$$

for the matrix S defined by the infinite geometric series

$$S = E + A + A^2 + \cdots = \sum_{\nu=0}^{\infty} A^\nu.$$

Under what circumstances an infinite geometric series of matrices or a *Neumann series*, as it is occasionally called, converges will be investigated in the next section.

Matrix polynomials may be handled very much like ordinary polynomials in x. For example, an identity between two polynomials in x implies the corresponding identity for an arbitrary matrix A. Thus the identity

$$x^3 + 2x^2 + 3x + 4 \equiv (x^2 + 1)(x + 2) + (2x + 2)$$

corresponds to the relation

$$A^3 + 2A^2 + 3A + 4E \equiv (A^2 + E)(A + 2E) + (2A + 2E)$$

valid for every matrix A. The factorization

$$f(x) = a_0 + a_1x + \cdots + a_mx^m = a_m(x - x_1)(x - x_2) \cdots (x - x_m),$$

where $x_1, x_2, \cdots, x_m$ are the zeros of the polynomial $f(x)$, leads to the matrix equation

$$\begin{aligned} f(A) &= a_0E + a_1A + \cdots + a_mA^m \\ &= a_m(A - x_1E)(A - x_2E) \cdots (A - x_mE) \end{aligned}$$

for every matrix A.

Every matrix A with components a_{ik}, which may in general be complex, is associated with certain other matrices. If $\bar{a}_{ik}$ is the complex number conjugate to a_{ik}, then the matrix $\bar{A} = (\bar{a}_{ik})$ is called the *conjugate* matrix; the matrix $A' = (a_{ki})$ obtained by interchanging corresponding rows and columns of A is called the *transposed* matrix or the *transpose* of A and $A^* = \bar{A}' = (\bar{a}_{ki})$ the *conjugate transpose* of A. The conjugate transpose is thus obtained by replacing the elements by their complex conjugates and interchanging rows and columns.

The equation

$$(AB)' = B'A'$$

is immediately verifiable. A matrix for which $A = A'$ is called *symmetric*; a real matrix which satisfies

$$AA' = E$$

is called *orthogonal*. Finally, a complex matrix is called *unitary* if it satisfies

$$AA^* = E.$$

The inversion of the linear transformation (7) is possible for arbitrary y_i, as is known from the theory of determinants, if and only if the determinant $\mathsf{A} = |a_{ik}|$ does not vanish. In this case the solution is uniquely determined and is given by a corresponding transformation

$$x_i = \sum_{k=1}^{n} \check{a}_{ik} y_k \qquad (i = 1, \cdots, n). \tag{11}$$

The coefficients $\check{a}_{ik}$ are given by

$$\check{a}_{ik} = \frac{\mathsf{A}_{ki}}{\mathsf{A}} \tag{12}$$

where A_{ki} is the cofactor to the element a_{ki} in the matrix A. The matrix $\check{A} = (\check{a}_{ik})$ is called the *reciprocal* or *inverse of* A and is distinguished by the fact that it satisfies

$$A\check{A} = \check{A}A = E.$$

We denote this uniquely determined matrix by A^{-1} instead of $\check{A}$; the determinant of A^{-1} has the value A^{-1}. Thus the solution of a system of equations whose matrix A has a nonvanishing determinant is characterized, in the language of matrix algebra, by a matrix $B = A^{-1}$ which satisfies the relations

$$AB = BA = E.$$

4. Bilinear, Quadratic, and Hermitian Forms. To write the linear equations (7) concisely we may employ the *bilinear form* which corresponds to the matrix A. This bilinear form

$$A(u, x) = \sum_{i,k=1}^{n} a_{ik} u_i x_k \tag{13}$$

is obtained by multiplying the linear forms in $x_1, x_2, \cdots, x_n$ or the left-hand side in equation (7) by undetermined quantities $u_1, u_2, \cdots, u_n$ and adding. In this way we obtain from the system of equations (7) the single equation

$$A(u, x) = E(u, y) \tag{14}$$

valid for all u; here $E(u, y) = \sum_{i=1}^{n} u_i y_i$ is the bilinear form corresponding to the unit matrix, the *unit bilinear form*. The *symbolic product* of two bilinear forms $A(u, x)$ and $B(u, x)$ with matrices A and B is defined as the bilinear form $C(u, x)$ with the matrix $C = AB$; the h-th power $A^h(u, x)$ is often called the *h-fold iterated form*. The "*reciprocal bilinear form*" $A^{-1}(u, x)$ with the matrix A^{-1} may, according to the theory of determinants, be written in the form

$$A^{-1}(u, x) = -\frac{\mathsf{A}(u, x)}{\mathsf{A}}, \tag{15}$$

where

$$\mathsf{A}(u, x) = \begin{vmatrix} 0 & u_1 & \cdots & u_n \\ x_1 & a_{11} & \cdots & a_{1n} \\ \cdots & \cdots & \cdots & \cdots \\ x_n & a_{n1} & \cdots & a_{nn} \end{vmatrix} = -\sum_{i,k=1}^{n} \mathsf{A}_{ik} x_i u_k .$$

The *symmetric linear transformations*, characterized by the condition $a_{ik} = a_{ki}$, are of special interest. To investigate them it is sufficient to consider the *quadratic form*

$$A(x, x) = \sum_{i,k=1}^{n} a_{ik} x_i x_k \qquad (a_{ki} = a_{ik})$$

which is obtained from the bilinear form by putting $u_i = x_i$. For, from a quadratic form $A(x, x)$ one can obtain a symmetric bilinear form

$$\sum_{i,k=1}^{n} a_{ik} u_i x_k = \tfrac{1}{2} \sum_{i=1}^{n} u_i \frac{\partial A(x, x)}{\partial x_i} = \frac{A(x + u, x + u) - A(x, x) - A(u, u)}{2},$$

which is called the *polar form* corresponding to the quadratic form $A(x, x)$.

If $A(u, x) = \sum_{i,k=1}^{n} a_{ik} u_i x_k$ is an arbitrary nonsymmetric bilinear form (with real coefficients), then $AA'(u, x)$ and $A'A(u, x)$ are always symmetric bilinear forms; specifically we have

$$AA'(u, x) = \sum_{k=1}^{n} \left(\sum_{i=1}^{n} a_{ik} x_i \sum_{j=1}^{n} a_{jk} u_j \right)$$

$$A'A(u, x) = \sum_{i=1}^{n} \left(\sum_{k=1}^{n} a_{ik} x_k \sum_{j=1}^{n} a_{ij} u_j \right).$$

The corresponding quadratic forms

$$AA'(x, x) = \sum_{k=1}^{n} \left(\sum_{i=1}^{n} a_{ik} x_i \right)^2,$$

$$A'A(x, x) = \sum_{i=1}^{n} \left(\sum_{k=1}^{n} a_{ik} x_k \right)^2,$$

which are sums of squares, assume only non-negative values. Forms of this kind are called *positive definite* quadratic forms.

An important generalization of the quadratic form is the *Hermitian form*. A Hermitian form is a bilinear form

$$A(u, x) = \sum_{i,k=1}^{n} a_{ik} u_i x_k$$

whose coefficients a_{ik} have complex values subject to the condition

$$a_{ik} = \bar{a}_{ki}.$$

Thus a Hermitian form assumes real values if the variables u_i are taken to be the complex conjugates of x_i; it is usually written in the form

$$H(x, \bar{x}) = \sum_{i,k=1}^{n} a_{ik} x_i \bar{x}_k = \sum_{i,k=1}^{n} a_{ki} \bar{x}_i x_k.$$

To an arbitrary bilinear form

$$A(u, x) = \sum_{i,k=1}^{n} a_{ik} u_i x_k$$

with complex coefficients there correspond the Hermitian forms

$$AA^*(x, \bar{x}) = A\bar{A}'(x, \bar{x}) = \sum_{k=1}^{n} \left| \sum_{i=1}^{n} a_{ik} x_i \right|^2$$

and

$$A^*A(x, \bar{x}) = \bar{A}'A(x, \bar{x}) = \sum_{i=1}^{n} \left| \sum_{k=1}^{n} a_{ik} \bar{x}_k \right|^2.$$

If the variables of a bilinear form

$$A(x, y) = \sum_{i,k=1}^{n} a_{ik} x_i y_k$$

are subjected to the two transformations

$$x_i = \sum_{j=1}^{n} c_{ij}\zeta_j \qquad \text{and} \qquad y_k = \sum_{l=1}^{n} b_{kl}\eta_l$$

with matrices C and B, respectively, we obtain

$$A(x, y) = \sum_{i,k=1}^{n} a_{ik}x_i y_k = \sum_{i,j,k,l=1}^{n} a_{ik}c_{ij}b_{kl}\zeta_j\eta_l$$

$$= \sum_{j,l=1}^{n} p_{jl}\zeta_j\eta_l\,; \qquad p_{jl} = \sum_{i,k=1}^{n} a_{ik}c_{ij}b_{kl}\,.$$

Thus A is transformed into a bilinear form with the matrix

$$(p_{jl}) = C'AB,$$

whose determinant is, according to the theorem on the multiplication of determinants, equal to $\mathsf{AB\Gamma}$. In particular, if A is a quadratic form

$$K(x, x) = \sum_{p,q=1}^{n} k_{pq}x_p x_q$$

with the symmetric matrix $K = (k_{pq})$ and the determinant $\mathsf{K} = |k_{pq}|$, and if we set $C = B$, and transform the variables x we obtain a quadratic form with the symmetric matrix $C'KC$ whose determinant is $\mathsf{K\Gamma}^2$.

5. Orthogonal and Unitary Transformations. We now consider the problem of finding "*orthogonal*" linear transformations L

$$x_p = \sum_{q=1}^{n} l_{pq}y_q = L_p(y) \qquad (p = 1, \cdots, n), \tag{16}$$

with the real matrix $L = (l_{pq})$ and the determinant $\Lambda = |l_{pq}|$, i.e. transformations which transform the *unit quadratic form*

$$E(x, x) = \sum_{p=1}^{n} x_p^2$$

into itself, thus satisfying the relation

$$E(x, x) = E(y, y) \tag{17}$$

for arbitrary y.

Applying our rules of transformation to the quadratic form

$A(x, x) = E(x, x)$, we find that requirement (17) yields the equations

$$L'EL = L'L = LL' = E; \qquad L' = L^{-1} \tag{18}$$

as a necessary and sufficient condition for the orthogonality of L. Thus the transposed matrix of an orthogonal transformation is identical with its reciprocal matrix; therefore the solution of equations (16) is given by the transformation

$$y_p = \sum_{q=1}^{n} l_{qp}x_q = L'_p(x), \tag{19}$$

which is likewise orthogonal. We see that an orthogonal transformation is one whose matrix is orthogonal as defined in subsection 3. Written out in detail, the orthogonality conditions become

$$\sum_{\nu=1}^{n} l_{\nu p}^2 = 1, \qquad \sum_{\nu=1}^{n} l_{\nu p}l_{\nu q} = 0 \qquad (p \neq q) \tag{20}$$

or, equivalently,

$$\sum_{\nu=1}^{n} l_{p\nu}^2 = 1, \qquad \sum_{\nu=1}^{n} l_{p\nu}l_{q\nu} = 0 \qquad (p \neq q). \tag{21}$$

To express an orthogonal transformation in vector notation we prescribe a system of n orthogonal unit vectors $\mathfrak{l}_1, \mathfrak{l}_2, \cdots, \mathfrak{l}_n$ into which the coordinate vectors $\mathfrak{e}_1, \mathfrak{e}_2, \cdots, \mathfrak{e}_n$ are to be transformed. Then the vector $\mathfrak{x}$ is represented by

$$\mathfrak{x} = x_1\mathfrak{e}_1 + x_2\mathfrak{e}_2 + \cdots + x_n\mathfrak{e}_n = y_1\mathfrak{l}_1 + y_2\mathfrak{l}_2 + \cdots + y_n\mathfrak{l}_n .$$

Multiplying by $\mathfrak{e}_p$ we obtain $x_p = \sum_{q=1}^{n} y_q(\mathfrak{e}_p\mathfrak{l}_q)$; hence

$$l_{pq} = \mathfrak{e}_p \cdot \mathfrak{l}_q .$$

From (18) it follows that $\Delta^2 = 1$, i.e. that the determinant of an orthogonal transformation is either $+1$ or -1. Therefore the determinant of an arbitrary quadratic form is invariant with respect to orthogonal transformations.

Furthermore, the relation $L'(AB)L = (L'AL)(L'BL)$ follows from (18) for the matrices A, B, and L of any two bilinear forms and any orthogonal transformation. This means that the symbolic product of a number of bilinear forms may be transformed orthogonally by

subjecting each factor to the same orthogonal transformation. In particular, it follows that the orthogonal transforms of two reciprocal forms are also reciprocal.

The generalization of these considerations to Hermitian forms

$$H(x, \bar{x}) = \sum_{p,q=1}^{n} h_{pq} x_p \bar{x}_q$$

leads to *unitary transformations*. A unitary transformation

$$x_p = \sum_{q=1}^{n} l_{pq} y_q \qquad (p = 1, \cdots n)$$

is defined as a transformation (with complex coefficients l_{pq}) which transforms the *unit Hermitian form*

$$\sum_{p=1}^{n} | x_p |^2 = \sum_{p=1}^{n} x_p \bar{x}_p$$

into itself, i.e. for which

$$\sum_{p=1}^{n} | x_p |^2 = \sum_{p=1}^{n} | y_p |^2.$$

In exactly the same way as above one obtains the matrix equation

$$LL^* = L^*L = E$$

as a necessary and sufficient condition for the unitary character of the transformation whose matrix is L. Here $L^* = \bar{L}'$ is the conjugate transpose of L. L must therefore be a unitary matrix as defined in subsection 3. Specifically, a transformation is unitary if the following conditions hold:

$$(22) \qquad \sum_{\nu=1}^{n} | l_{\nu p} |^2 = 1, \qquad \sum_{\nu=1}^{n} l_{\nu p} l_{\nu q} = 0 \qquad (p \neq q),$$

or, equivalently,

$$(23) \qquad \sum_{\nu=1}^{n} | l_{p\nu} |^2 = 1, \qquad \sum_{\nu=1}^{n} l_{p\nu} l_{q\nu} = 0 \qquad (p \neq q).$$

The determinant of a unitary transformation has the absolute value 1, as follows immediately from the equation $LL^* = E$.

§2. *Linear Transformations with a Linear Parameter*

In many problems the system of equations of a linear transformation takes the form

$$x_i - \lambda \sum_{k=1}^{n} t_{ik} x_k = y_i \qquad (i = 1, \cdots, n) \tag{24}$$

where λ is a parameter (in general complex). The corresponding bilinear form is $E(u, x) - \lambda T(u, x)$, where $T(u, x)$ is the form whose matrix is (t_{ik}). As we have seen in the preceding section, the problem of solving the system of equations (24) is equivalent to the problem of finding the reciprocal bilinear form $R(u, y; \lambda)$ with the matrix R which satisfies the equation $(E - \lambda T)R = E$. We know that this reciprocal matrix R exists if and only if the determinant $| E - \lambda T |$ is different from zero.

Let us consider the zeros of the determinant $| E - \lambda T |$ or, equivalently, for $\kappa = 1/\lambda \neq 0$, the zeros of the determinant $| T - \kappa E |$. Clearly, $| T - \kappa E |$ is a polynomial in κ of the n-th degree. Therefore there exist n values of κ (namely the zeros of the polynomial) for which the form $R(u, y; \lambda)$ fails to exist. These values κ_i are known as the "*characteristic values*," "*proper values*," or "*eigenvalues*" of T with respect to the unit matrix E; they form the so-called "*spectrum*" of the matrix T.[1]

The particular form of equations (24) suggests a solution by iteration: In the equation

$$x_i = y_i + \lambda \sum_{k=1}^{n} t_{ik} x_k$$

we substitute for the quantities x_k on the right the expressions

$$y_k + \lambda \sum_{j=1}^{n} t_{kj} x_j ,$$

and then again repeat this substitution. The procedure is conveniently described if we write $R = E + \lambda TR$ and continue:

$$\begin{aligned} R &= E + \lambda TR = E + \lambda T + \lambda^2 T^2 R \\ &= E + \lambda T + \lambda^2 T^2 + \lambda^3 T^3 R = \cdots . \end{aligned}$$

[1] Sometimes the set of values $\lambda_i = 1/\kappa_i$, for which no reciprocal of $E - \lambda T$ exists, is called the spectrum. We shall call this the "reciprocal spectrum" and the λ_i the "reciprocal eigenvalues."

We thus obtain an expression for R as an infinite series

$$R = E + \lambda T + \lambda^2 T^2 + \lambda^3 T^3 + \cdots,$$

which—assuming that it converges—represents the reciprocal matrix of $E - \lambda T$. To see this we simply multiply the series by $E - \lambda T$ and remember that symbolic multiplication may be carried out term by term provided the result converges. It is then immediately clear that the representation

$$R = (E - \lambda T)^{-1} = E + \lambda T + \lambda^2 T^2 + \lambda^3 T^3 + \cdots$$

is, formally, completely equivalent to the ordinary geometric series. (Compare the discussion of geometric series on page 9, where we need only put $A = \lambda T$ to obtain equivalence.)

Let us now represent our original system of equations using bilinear forms instead of the corresponding matrices:

$$E(u, x) - \lambda T(u, x) = E(u, y).$$

We may write the solution of this equation in the form

$$E(u, y) + \lambda \mathsf{T}(u, y; \lambda) = E(u, x),$$

which is completely symmetric to it; here

$$\begin{aligned} \mathsf{T}(u, y; \lambda) &= T + \lambda T^2 + \lambda^2 T^3 + \cdots \\ &= \frac{R(u, y; \lambda) - E(u, y)}{\lambda}. \end{aligned}$$

The form T is called the *resolvent* of T.

The convergence of the above *Neumann series* for R or T for sufficiently small $|\lambda|$ is easily proved: If M is an upper bound of the absolute values of the numbers t_{ik}, it follows immediately that upper bounds for the absolute values of the coefficients of the forms T^2, T^3, $\cdots$, T^h are given by nM^2, n^2M^3, $\cdots$, $n^{h-1}M^h$. Thus

$$(M + \lambda n M^2 + \lambda^2 n^2 M^3 + \cdots) \cdot (|u_1| + |u_2| + \cdots + |u_n|)(|y_1| + \cdots + |y_n|)$$

is a majorant of the Neumann series for $\mathsf{T}(u, y; \lambda)$; it is certainly convergent for $|\lambda| < 1/nM$. Therefore our Neumann series also

converges for sufficiently small $|\lambda|$ and actually represents the resolvent of $T(u, x)$.[1]

The above estimate proves, incidentally, that in any everywhere convergent power series $f(x) = \sum_{\nu=0}^{\infty} c_\nu x^\nu$ we may replace x by an arbitrary matrix A and obtain a new matrix $f(A) = \sum_{\nu=0}^{\infty} c_\nu A^\nu$. Thus, in particular, the matrix e^A always exists.

While the above expression for R or T converges only for sufficiently small $|\lambda|$, we may obtain from equation (15) of the previous section an expression for the reciprocal form or matrix $R = (E - \lambda T)^{-1}$ which retains its meaning even outside the region of convergence. For, if we identify the form $E - \lambda T$ with the form $A(u, x)$, we immediately obtain, for the reciprocal form,

$$R(u, y; \lambda) = -\frac{\Delta(u, y; \lambda)}{\Delta(\lambda)}$$

[1] The convergence of the majorant obtained above evidently becomes worse with increasing n. It may, however, be pointed out that, by slightly refining the argument, an upper bound for the coefficients of the form T can be obtained which is independent of n and which, therefore, can be used for the generalization to infinitely many variables. We denote the elements of the matrix T^ν by $t_{ik}^{(\nu)}$ and set

$$\sum_{\alpha=1}^{n} |t_{p\alpha}^{(1)}| = z_p.$$

Then, if $\bar{M}$ is an upper bound for all the n quantities z_p, it follows, as will be shown below by induction, that

$$\sum_{q=1}^{n} |t_{pq}^{(\nu)}| \leq \bar{M}^\nu;$$

therefore,

$$|t_{pq}^{(\nu)}| \leq \bar{M}^\nu$$

for $p, q = 1, 2, \cdots, n$ and every ν. From this we see that our Neumann series converges for $|\lambda| < 1/\bar{M}$. We thus have a bound which does not depend on n explicitly.

To prove the above inequality for arbitrary ν we assume it to be proved for the index $\nu - 1$; we then have

$$\sum_{q=1}^{n} |t_{pq}^{(\nu)}| = \sum_{q=1}^{n} \left| \sum_{\alpha=1}^{n} t_{p\alpha}^{(1)} t_{\alpha q}^{(\nu-1)} \right| \leq \sum_{q=1}^{n} \sum_{\alpha=1}^{n} |t_{p\alpha}^{(1)}| \, |t_{\alpha q}^{(\nu-1)}|$$

$$= \sum_{\alpha=1}^{n} |t_{p\alpha}^{(1)}| \left(\sum_{q=1}^{n} |t_{\alpha q}^{(\nu-1)}| \right) \leq \bar{M}^{\nu-1} \cdot \sum_{\alpha=1}^{n} |t_{p\alpha}^{(1)}| \leq \bar{M}^\nu.$$

Since the inequality is valid for $\nu = 1$, it is proved for arbitrary ν.

and, for the resolvent T,

$$\mathsf{T}(u, y; \lambda) = -\frac{\Delta(u, y; \lambda)}{\lambda\Delta(\lambda)} - \frac{1}{\lambda} E(u, y),$$

where

$$\Delta(u, y; \lambda) = \begin{vmatrix} 0 & u_1 & \cdots & u_n \\ y_1 & 1 - \lambda t_{11} & \cdots & -\lambda t_{1n} \\ \cdots & \cdots & \cdots & \cdots \\ y_n & -\lambda t_{n1} & \cdots & 1 - \lambda t_{nn} \end{vmatrix}$$

and

$$\Delta(\lambda) = \begin{vmatrix} 1 - \lambda t_{11} & -\lambda t_{12} & \cdots & -\lambda t_{1n} \\ -\lambda t_{21} & 1 - \lambda t_{22} & \cdots & -\lambda t_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ -\lambda t_{n1} & -\lambda t_{n2} & \cdots & 1 - \lambda t_{nn} \end{vmatrix}$$

are polynomials in λ of at most the $(n - 1)$-st and n-th degree. Thus the zeros of $\Delta(\lambda)$ form the reciprocal spectrum of the form T as defined above, i.e. the totality of values of λ for which the form $E - \lambda T$ has no reciprocal.

By means of the formula

$$T + \lambda T^2 + \lambda^2 T^3 + \cdots = -\frac{\Delta(u, y; \lambda)}{\lambda\Delta(\lambda)} - \frac{1}{\lambda} E(u, y)$$

the series on the left, which does not converge for all λ, is continued analytically into the entire λ-plane. The reciprocal form R, as well as the resolvent T, is a rational function of λ whose poles are given by the spectrum of the form T.

If we expand the determinants $\Delta(u, y; \lambda)$ and $\Delta(\lambda)$ in powers of λ, we obtain the expressions

$$\Delta(u, y; \lambda) = \Delta_1(u, y) - \lambda\Delta_2(u, y) + \lambda^2\Delta_3(u, y) - \cdots + (-1)^n\lambda^{n-1}\Delta_n(u, y),$$

$$\Delta(\lambda) = 1 - \lambda\Delta_1 + \lambda^2\Delta_2 - \cdots + (-1)^n\lambda^n\Delta_n,$$

where

$$\Delta_h(u, y) = \sum \begin{vmatrix} 0 & u_{p_1} & \cdots & u_{p_h} \\ y_{p_1} & t_{p_1 p_1} & \cdots & t_{p_1 p_h} \\ \cdots & \cdots & \cdots & \cdots \\ y_{p_h} & t_{p_h p_1} & \cdots & t_{p_h p_h} \end{vmatrix}$$

and

$$\Delta_h = \sum \begin{vmatrix} t_{p_1 p_1} & t_{p_1 p_2} & \cdots & t_{p_1 p_h} \\ t_{p_2 p_1} & t_{p_2 p_2} & \cdots & t_{p_2 p_h} \\ \cdots & \cdots & \cdots & \cdots \\ t_{p_h p_1} & t_{p_h p_2} & \cdots & t_{p_h p_h} \end{vmatrix}.$$

The summations here are extended over all integers $p_1, p_2, \cdots, p_h$ from 1 to n with $p_1 < p_2 < \cdots < p_h$.

It is often advantageous to consider the form $\kappa E - T$ with the determinant

$$\begin{vmatrix} \kappa - t_{11} & -t_{12} & \cdots & -t_{1n} \\ -t_{21} & \kappa - t_{22} & \cdots & -t_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ -t_{n1} & -t_{n2} & \cdots & \kappa - t_{nn} \end{vmatrix} = \varphi(\kappa).$$

Its zeros $\kappa_1, \kappa_2, \cdots, \kappa_n$ (eigenvalues of T) are the reciprocals of the zeros of $\Delta(\lambda)$. For the reciprocal form $(\kappa E - T)^{-1}$, which exists for all values of κ different from $\kappa_1, \kappa_2, \cdots, \kappa_n$, one obtains the Neumann series expansion

$$(\kappa E - T)^{-1} = \frac{E}{\kappa} + \frac{T}{\kappa^2} + \frac{T^2}{\kappa^3} + \cdots$$

which is valid for sufficiently large values of $|\kappa|$. A noteworthy conclusion can be drawn from this expansion. It is clear from the above discussion that the left side is a rational function of κ with the denominator $\varphi(\kappa)$; therefore $\varphi(\kappa)(\kappa E - T)^{-1}$ must be a form which is integral and rational in κ and its expansion in powers of κ can contain no negative powers. Accordingly, if we multiply the above equation by $\varphi(\kappa) = \kappa^n + c_1\kappa^{n-1} + \cdots + c_n$, all the coefficients of negative powers of κ in the resulting expression on the right must

vanish. But the coefficient of κ^{-1} is, as is seen immediately, the expression $T^n + c_1T^{n-1} + \cdots + c_n$, and we thus arrive at the following theorem, which is due to Cayley: *If the determinant of* $\kappa E - T$ *is denoted by* $\varphi(\kappa)$, *then the relation*

$$\varphi(T) = 0$$

is satisfied by the matrix T.

Another important aspect of the spectrum of the eigenvalues $\kappa_1, \kappa_2, \cdots, \kappa_n$ is expressed by the following theorem:

If the eigenvalues of a matrix T *are* $\kappa_1, \kappa_2, \cdots, \kappa_n$ *and if* $g(x)$ *is any polynomial in* x, *then the eigenvalues of the matrix* $g(T)$ *are* $g(\kappa_1)$, $g(\kappa_2), \cdots, g(\kappa_n)$.

To prove this we start from the relation

$$|\kappa E - T| = \varphi(\kappa) = \prod_{\nu=1}^{n} (\kappa - \kappa_\nu),$$

which is an identity in T. We wish to obtain the relation

$$|\kappa E - g(T)| = \prod_{\nu=1}^{n} (\kappa - g(\kappa_\nu)).$$

Let $h(x)$ be an arbitrary polynomial of degree r which may be written in terms of its zeros $x_1, x_2, \cdots, x_n$ in the form

$$h(x) = a \prod_{\rho=1}^{r} (x - x_\rho).$$

Then the identity

$$h(T) = a \prod_{\rho=1}^{r} (T - x_\rho E)$$

holds for an arbitrary matrix T. By considering the determinants of the matrices in this equation we obtain

$$\begin{aligned} |h(T)| &= a^n \prod_{\rho=1}^{r} |T - x_\rho E| = (-1)^{nr} a^n \prod_{\rho=1}^{r} |x_\rho E - T| \\ &= (-1)^{nr} a^n \prod_{\rho=1}^{r} \varphi(x_\rho) = (-1)^{nr} a^n \prod_{\rho=1}^{r} \left(\prod_{\nu=1}^{n} (x_\rho - \kappa_\nu) \right) \\ &= (-1)^{nr} (-1)^{nr} a^n \prod_{\nu=1}^{n} \left(\prod_{\rho=1}^{r} (\kappa_\nu - x_\rho) \right) = \prod_{\nu=1}^{n} h(\kappa_\nu). \end{aligned}$$

If we now let $h(T)$ be the function $\kappa E - g(T)$, the desired equation

$$| \kappa E - g(T) | = \prod_{\nu=1}^{n} (\kappa - g(\kappa_\nu))$$

follows immediately.

§3. *Transformation to Principal Axes of Quadratic and Hermitian Forms*

Linear transformations $x = Z(y)$ which reduce a quadratic form

$$K(x, x) = \sum_{p,q=1}^{n} k_{pq} x_p x_q$$

to a linear combination of squares

$$K(x, x) = \sum_{p=1}^{n} \kappa_p y_p^2$$

are highly important in algebra. We are particularly interested in reducing $K(x, x)$ to this form by means of an orthogonal transformation

$$x_p = \sum_{q=1}^{n} l_{qp} y_q = L_p(y) \qquad (p = 1, \cdots, n).$$

Transformations of this kind are called *transformations to principal axes.*

1. Transformation to Principal Axes on the Basis of a Maximum Principle. Let us first convince ourselves that a transformation to principal axes is always possible for any given quadratic form $K(x, x)$. To do this we use the theorem that a continuous function of several variables (which are restricted to a finite closed domain) assumes a greatest value somewhere in this domain (Theorem of Weierstrass).[1]

[1] The transformation to principal axes may also be accomplished by direct algebraic methods. An orthogonal matrix L is required, such that $L'KL = D$ is a diagonal matrix with diagonal elements $\kappa_1, \kappa_2, \cdots, \kappa_n$. From the relation $KL = LD$ we obtain the equations

$$\sum_{q=1}^{n} k_{pq} l_{qi} = l_{pi} \kappa_i$$

for the matrix elements l_{qi}, which yield the κ_i as roots of equation (30), cf. p. 27. Then, on the basis of simple algebraic considerations we can construct an orthogonal system of n^2 quantities l_{qi}. The method used in the text is preferable to the algebraic method in that it may be generalized to a larger class of transcendental problems.

According to this theorem, there exists a unit vector $\mathfrak{l}_1$ with components $l_{11}, l_{12}, \cdots, l_{1n}$ such that, for $x_1 = l_{11}, \cdots, x_n = l_{1n}$, $K(x, x)$ assumes its greatest value, say κ_1, subject to the subsidiary condition

$$\sum_{p=1}^{n} x_p^2 = 1. \tag{25}$$

Geometrically, the vector $\mathfrak{l}_1$ represents on the "unit sphere" (25) a point P so that the surface of the second degree $K(x, x) = \text{const.}$ containing P touches the unit sphere at P.

There exists, moreover, a unit vector $\mathfrak{l}_2$, orthogonal to $\mathfrak{l}_1$, with components $l_{21}, \cdots, l_{2n}$ such that, for $x_1 = l_{21}, \cdots, x_n = l_{2n}$, $K(x, x)$ assumes the greatest possible value κ_2 subject to the condition

$$\sum_{p=1}^{n} l_{1p} x_p = 0 \tag{26}$$

in addition to condition (25). The problem solved by $\mathfrak{l}_1$ for the whole unit sphere is solved by $\mathfrak{l}_2$ for the manifold formed by the intersection of the unit sphere and the "plane" (26).

Furthermore, there exists a unit vector $\mathfrak{l}_3$, orthogonal to $\mathfrak{l}_1$ and $\mathfrak{l}_2$, with components $l_{31}, l_{32}, \cdots, l_{3n}$ such that, for $x_i = l_{3i}$ $(i = l, \cdots, n)$, $K(x, x)$ takes on its greatest value κ_3, subject to the subsidiary conditions (25), (26), and

$$\sum_{p=1}^{n} l_{2p} x_p = 0. \tag{27}$$

Continuing in this manner we obtain a system of n mutually orthogonal vectors $\mathfrak{l}_1, \mathfrak{l}_2, \cdots, \mathfrak{l}_n$, which will be called the "*principal axis vectors*" or "*eigenvectors.*" According to (21) their components l_{qp} define an orthogonal transformation

$$x_p = \sum_{q=1}^{n} l_{qp} y_q \qquad (p = 1, \cdots, n); \tag{28}$$

this transformation, we assert, is the solution of our problem.

Since equations (28) are solved by

$$y_p = \sum_{q=1}^{n} l_{pq} x_q \qquad (p = 1, \cdots, n), \tag{29}$$

the equation $\mathfrak{x} = \mathfrak{l}_p$ is equivalent to the statement $y_p = 1$, $y_q = 0$

for $q \neq p$. Thus, in particular, the maximum κ_1 is attained for $y_1 = 1, y_2 = 0, \cdots, y_n = 0$; hence, in the transformed form

$$C(y, y) = \sum_{p,q=1}^{n} c_{pq} y_p y_q = K(x, x)$$

the first coefficient c_{11} equals κ_1. The form

$$H(y, y) = \sum_{p,q=1}^{n} h_{pq} y_p y_q = C(y, y) - \kappa_1(y_1^2 + \cdots + y_n^2)$$

assumes, moreover, no positive values. For, by the maximum character of κ_1, $H(y, y)$ is nonpositive provided $\sum_{p=1}^{n} x_p^2 = \sum_{p=1}^{n} y_p^2 = 1$; hence it is nonpositive for all y_i with $\sum_{p=1}^{n} y_p^2 \neq 0$. If y_1 should occur in the expression for $H(y, y)$, e.g. if $h_{12} = h_{21}$ were different from zero, we would obtain the value

$$2h_{12}\epsilon + h_{22}\epsilon^2 = \epsilon(2h_{12} + h_{22}\epsilon)$$

for $H(y, y)$ with

$$y_1 = 1, \qquad y_2 = \epsilon, \qquad y_3 = \cdots = y_n = 0.$$

This could be made positive by a suitable choice of ϵ.

It has thus been shown that, after the transformation, $K(x, x)$ is reduced to

$$C(y, y) = \kappa_1 y_1^2 + C_1(y, y),$$

where $C_1(y, y)$ is a quadratic form in the $n - 1$ variables $y_2, y_3, \cdots, y_n$. If the subsidiary condition $y_1 = 0$ is imposed the transformed form is equal to $C_1(y, y)$. In the same way we may now conclude that $C_1(y, y)$ is of the form $\kappa_2 y_2^2 + C_2(y, y)$, where $C_2(y, y)$ depends only on the $n - 2$ variables $y_3, y_4, \cdots, y_n$, and so forth.

Thus we have demonstrated the possibility of a transformation to principal axes so that

$$\sum_{p,q=1}^{n} k_{pq} x_p x_q = \sum_{p=1}^{n} \kappa_p y_p^2, \qquad \sum_{p=1}^{n} x_p^2 = \sum_{p=1}^{n} y_p^2.$$

We might note that the corresponding minimum problem would have served equally well as a starting point for the proof; i.e. we might have looked for the minimum of $K(x, x)$, subject to the auxiliary

condition $E(x, x) = 1$. In that case we would have arrived at the quantities $\kappa_1, \kappa_2, \cdots, \kappa_n$ in the reverse order. One could also keep $K(x, x)$ constant and look for the maxima or minima of $E(x, x)$; then the minimum values λ_i would be the reciprocals of the κ_i.

2. Eigenvalues. We shall now show that the values κ_i defined in the previous subsection as successive maxima are identical with the eigenvalues as introduced in §2.

The equation

$$\varphi(\kappa) = (\kappa - \kappa_1)(\kappa - \kappa_2) \cdots (\kappa - \kappa_n) = 0$$

satisfied by the numbers κ_i, may be written in the form

$$\begin{vmatrix} \kappa - \kappa_1 & 0 & 0 & \cdots & 0 \\ 0 & \kappa - \kappa_2 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & \kappa - \kappa_n \end{vmatrix} = 0.$$

But this determinant is just the determinant of the quadratic form

$$\kappa \sum_{p=1}^{n} y_p^2 - \sum_{p=1}^{n} \kappa_p y_p^2,$$

which is obtained by applying an orthogonal transformation to the form

$$\kappa \sum_{p=1}^{n} x_p^2 - K(x, x).$$

Therefore the relation

$$\begin{vmatrix} \kappa - \kappa_1 & 0 & \cdots & 0 \\ 0 & \kappa - \kappa_2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \kappa - \kappa_n \end{vmatrix}$$

$$= \begin{vmatrix} \kappa - k_{11} & -k_{12} & \cdots & -k_{1n} \\ -k_{21} & \kappa - k_{22} & \cdots & -k_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ -k_{n1} & -k_{n2} & \cdots & \kappa - k_{nn} \end{vmatrix}$$

is an identity in κ. Consequently the numbers κ_i are the roots of the algebraic equation

$$(30) \qquad \begin{vmatrix} k_{11} - \kappa & k_{12} & \cdots & k_{1n} \\ k_{21} & k_{22} - \kappa & \cdots & k_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ k_{n1} & k_{n2} & \cdots & k_{nn} - \kappa \end{vmatrix} = 0$$

for the unknown κ; i.e. they are the eigenvalues introduced in §2.

Our method of derivation shows automatically that the roots of equation (30) are necessarily real if the k_{pq} are arbitrary real quantities subject to the condition $k_{pq} = k_{qp}$.[1] We may also remark in passing that the absolute values of the reciprocals of the eigenvalues are geometrically significant as the squares of the lengths of the principal axes of the surface $K(x, x) = 1$ in n-dimensional space. If at least one eigenvalue is equal to zero the form is said to be "*degenerate*"; it can then be represented as a form of less than n variables. It is clear from equation (30) that this is the case if and only if $| k_{pq} |$ vanishes. For $K(x, x)$ to be positive definite the condition $\kappa_p > 0$, $p = 1, 2, \cdots, n$ is necessary and sufficient.

Suppose the representation of a form $K(x, x)$ in terms of principal axes

$$K(x, x) = \sum_{p=1}^{n} \kappa_p y_p^2$$

is given. Then, using the properties of the orthogonal transformations of products discussed above, the expressions

$$K^2(x, x) = \sum_{p=1}^{n} \kappa_p^2 y_p^2, \qquad K^3(x, x) = \sum_{p=1}^{n} \kappa_p^3 y_p^2, \cdots$$

are easily obtained for the iterated forms. It follows that the eigenvalues of the h-fold iterated form $K^h(x, x)$ are the h-th powers of the eigenvalues of $K(x, x)$ (this also follows immediately from the theorem on page 22); moreover we see that, for even h, the form $K^h(x, x)$ is positive definite.

[1] Equation (30) is customarily called the secular equation because it occurs in the problem of secular perturbations of planetary orbits. For a direct proof that the eigenvalues are real, see Ch. III, §4, 2.

3. Generalization to Hermitian Forms. A transformation to principal axes can be carried out in exactly the same way for Hermitian forms. A Hermitian form

$$H(x, \bar{x}) = \sum_{p,q=1}^{n} h_{pq} x_p \bar{x}_q$$

with the matrix $H = \bar{H}'$ can always be transformed by a unitary transformation L, given by

$$x_p = \sum_{q=1}^{n} l_{qp} y_q ,$$

into the form

$$H(x, \bar{x}) = \sum_{p=1}^{n} \kappa_p y_p \bar{y}_p = \sum_{p=1}^{n} \kappa_p \mid y_p \mid^2 ,$$

where all the coefficients κ_p are real. These eigenvalues κ_m reappear as the maxima of the Hermitian form $H(x, \bar{x})$, subject to the auxiliary conditions

$$\sum_{p=1}^{n} \mid x_p \mid^2 = 1 \qquad \text{and} \qquad \sum_{p=1}^{n} l_{ip} \bar{x}_p = 0 \qquad (i = 1, \cdots, m - 1).$$

4. Inertial Theorem for Quadratic Forms. If we relinquish the requirement that the linear transformation be orthogonal, a quadratic form may be transformed into a sum of squares by many different transformations. In particular, after the above orthogonal transformation has been carried out, any transformation in which each variable is simply multiplied by a factor of proportionality leaves the character of the form as a sum of squares unaltered. Thus it is possible to transform the form in such a way that all the (real) coefficients have the value $+1$ or -1. The following theorem, known as the *inertial theorem for quadratic forms*, holds:

The number of positive and negative coefficients, respectively, in a quadratic form reduced to an expression $\sum c_p z_p^2$ *by means of a nonsingular real linear transformation does not depend on the particular transformation.*

Proof: The positive and negative coefficients may be made equal to $+1$ and -1, respectively. Suppose, now, that the quadratic form $K(x, x)$ is transformed by two different transformations into

$y_1^2 + \cdots + y_r^2 - y_{r+1}^2 - \cdots - y_n^2$ and $z_1^2 + \cdots + z_s^2 - z_{s+1}^2 - \cdots - z_n^2$ with $r < s$. We then have

$$y_1^2 + \cdots + y_r^2 + z_{s+1}^2 + \cdots + z_n^2 = y_{r+1}^2 + \cdots + y_n^2 + z_1^2 + \cdots + z_s^2 .$$

Let us consider the conditions $y_1 = \cdots = y_r = z_{s+1} = \cdots = z_n = 0$, which imply that the remaining y_i also vanish. By imagining the z_i expressed in terms of the y_i and regarding these conditions as a system of fewer than n equations in the y_i, we obtain the contradiction that there exists a non-vanishing solution vector.

5. Representation of the Resolvent of a Form. The resolvent of the quadratic form $K(x, x)$ can be expressed in a suggestive way. According to §2 the resolvent may be defined by the symbolic equation

$$\mathsf{K}(x, x; \lambda) = \frac{[E(x, x) - \lambda K(x, x)]^{-1} - E(x, x)}{\lambda}.$$

We suppose that $K(x, x)$ has been brought into the form

$$K(x, x) = \sum_{p=1}^{n} \frac{y_p^2}{\lambda_p}.$$

The resolvent of $\sum_{p=1}^{n} y_p^2/\lambda_p$ must be identical with the resolvent of $K(x, x)$, since $[E(x, x) - \lambda K(x, x)]^{-1}$ goes over into

$$\left[E(y, y) - \lambda \sum_{p=1}^{n} \frac{y_p^2}{\lambda_p}\right]^{-1}$$

when the transformation is applied. Now the following relations hold:

$$\frac{1}{\lambda}\left[\left(\sum_{p=1}^{n} y_p^2 - \lambda \sum_{p=1}^{n} \frac{y_p^2}{\lambda_p}\right)^{-1} - E(y, y)\right]$$

$$= \frac{1}{\lambda}\left[\left(\sum_{p=1}^{n} \frac{\lambda_p - \lambda}{\lambda_p} y_p^2\right)^{-1} - E(y, y)\right] = \frac{1}{\lambda}\left[\sum_{p=1}^{n} \frac{\lambda_p}{\lambda_p - \lambda} y_p^2 - E(y, y)\right]$$

$$= \frac{1}{\lambda}\left[\sum_{p=1}^{n} \frac{\lambda_p}{\lambda_p - \lambda} y_p^2 - \sum_{p=1}^{n} y_p^2\right] = \sum_{p=1}^{n} \frac{y_p^2}{\lambda_p - \lambda}.$$

If we now transform back to the variables x_p, using the notation (19) we obtain the expression

$$\mathsf{K}(x, x; \lambda) = \sum_{p=1}^{n} \frac{[L_p'(x)]^2}{\lambda_p - \lambda} \tag{31}$$

for the resolvent of $K(x, x)$; thus, for the bilinear form, we have

$$\mathsf{K}(u, x; \lambda) = \sum_{p=1}^{n} \frac{L_p'(u)L_p'(x)}{\lambda_p - \lambda}. \tag{32}$$

From this representation it is evident, incidentally, that the residue of the rational function $\mathsf{K}(u, x; \lambda)$ of λ at the point λ_p is equal to $-L_p'(u)L_p'(x)$, assuming that $\lambda_p \neq \lambda_q$ for $p \neq q$.

6. Solution of Systems of Linear Equations Associated with Forms. In conclusion we shall present, with the help of the eigenvectors, the solution of the system of linear equations

$$x_p - \lambda \sum_{q=1}^{n} k_{pq}x_q = y_p \qquad (p = 1, \cdots, n) \tag{33}$$

associated with the quadratic form

$$K(x, x) = \sum_{p,q=1}^{n} k_{pq}x_p x_q .$$

If we apply the transformation to principal axes

$$x_p = \sum_{q=1}^{n} l_{qp}u_q , \qquad y_p = \sum_{q=1}^{n} l_{qp}v_q$$

to the variables x_i and y_i, $K(x, x)$ goes over into

$$\sum_{q=1}^{n} \kappa_q u_q^2 ,$$

and the bilinear form $K(x, z)$ is similarly transformed. Hence, our system of equations (33) becomes

$$u_p - \lambda\kappa_p u_p = v_p \qquad (p = 1, \cdots, n). \tag{34}$$

the solution of which is

$$u_p = \frac{v_p}{1 - \lambda\kappa_p} = \frac{v_p}{1 - \dfrac{\lambda}{\lambda_p}} = \frac{\lambda_p}{\lambda_p - \lambda} v_p . \tag{35}$$

In terms of the original variables, we obtain the equivalent formula for the solution

$$\mathbf{x} = \sum_{p=1}^{n} \frac{\mathbf{y} \cdot \mathbf{l}_p}{1 - \dfrac{\lambda}{\lambda_p}} \mathbf{l}_p , \tag{36}$$

in which the solution appears as a development in terms of the eigenvectors $\mathfrak{l}_1, \mathfrak{l}_2, \cdots, \mathfrak{l}_n$ of the form $K(x, x)$. We have here used the notation $\mathfrak{y} \cdot \mathfrak{l}_p = \sum_{q=1}^{n} l_{pq} y_q$.

The principal axis vector or eigenvector $\mathfrak{l}_p$ is itself the normalized solution of the homogeneous equations

$$x_q - \lambda_p \sum_{r=1}^{n} k_{qr} x_r = 0$$

or

$$u_q - \lambda_p \kappa_q u_q = 0 \qquad (q = 1, \cdots, n).$$

If, for $q \neq p$, all the κ_q are different from $\kappa_p = 1/\lambda_p$, there exists only one normalized solution,

$$u_p = 1,$$

$$u_q = 0 \qquad (q \neq p)$$

or

$$\mathfrak{x} = \mathfrak{l}_p.$$

If several characteristic numbers coincide the principal axis vectors are not uniquely determined.

§4. *Minimum-Maximum Property of Eigenvalues*

1. Characterization of Eigenvalues by a Minimum-Maximum Problem. In the above discussion we have obtained the eigenvalues by solving a series of maximum problems, each one of which depended on the solutions of the previous problems of the series. We shall now show that each eigenvalue can be directly characterized as the solution of a somewhat different problem in which all reference to the solutions of previous problems is avoided.

The problem is to maximize the form

$$K(x, x) = \sum_{p,q=1}^{n} k_{pq} x_p x_q$$

if the condition (25)

$$\sum_{p=1}^{n} x_p^2 = 1$$

is imposed and if the $h - 1$ equations

$$\sum_{p=1}^{n} \alpha_{\nu p} x_p = 0 \qquad (\nu = 1, \cdots, h-1; h \leq n) \tag{37}$$

must be satisfied. This maximum value of $K(x, x)$ is of course a function of the parameters $\alpha_{\nu p}$. We now choose the $\alpha_{\nu p}$ in such a way as to give this maximum its least possible value. We assert that this minimum value of the maximum is just the h-th eigenvalue κ_h of $K(x, x)$, provided the eigenvalues are ordered in a sequence of decreasing values, κ_1 being the greatest eigenvalue, κ_2 the next, and so on.

The transformation to principal axes changes $K(x, x)$ into

$$\sum_{p=1}^{n} \kappa_p y_p^2 \qquad (\kappa_1 \geq \cdots \geq \kappa_n),$$

condition (25) into

$$\sum_{p=1}^{n} y_p^2 = 1, \tag{38}$$

and equations (37) into

$$\sum_{p=1}^{n} \beta_{\nu p} y_p = 0, \qquad (\nu = 1, \cdots h-1; h \leq n) \tag{39}$$

where the $\beta_{\nu p}$ are new parameters. If we set

$$y_{h+1} = \cdots = y_n = 0$$

equations (39) become $h - 1$ equations in h unknowns $y_1, y_2, \cdots, y_h$, which can certainly be satisfied for a set of values y_i also satisfying (38). For these values we have

$$K(x, x) = \kappa_1 y_1^2 + \cdots + \kappa_h y_h^2 \geq \kappa_h (y_1^2 + \cdots + y_n^2) = \kappa_h .$$

Thus the required maximum of $K(x, x)$ for any set of values $\beta_{\nu p}$ is not less than κ_h; but it is just equal to κ_h if we take for (39) the equations

$$y_1 = \cdots = y_{h-1} = 0.$$

It follows therefore that:

The h-th eigenvalue κ_h of the quadratic form $K(x, x)$ is the least value which the maximum of $K(x, x)$ can assume if, in addition to the condition

$$\sum_{p=1}^{n} x_p^2 = 1,$$

$h - 1$ arbitrary linear homogeneous equations connecting the x_p are prescribed.

2. Applications. Constraints. This independent minimum-maximum property of the eigenvalues shows how the eigenvalues are changed if j independent constraints

$$\sum_{p=1}^{n} \gamma_{sp} x_p \qquad (s = 1, \cdots, j) \tag{40}$$

are imposed on the variables, so that $K(x, x)$ reduces to a quadratic form $\bar{K}(x, x)$ of $n - j$ independent variables. The h-th eigenvalue $\bar{\kappa}_h$ is obtained from the same minimum-maximum problem as κ_h, in which the totality of sets of admissible values x_i has been narrowed down by (40). Therefore the maximum, and thus the eigenvalue of $\bar{K}(x, x)$, certainly does not exceed the corresponding quantity for $K(x, x)$.

Furthermore, κ_{j+h} is the least maximum which $K(x, x)$ can possess if, in addition to (25), $h + j - 1$ linear homogeneous conditions are imposed on the x_p; κ_{j+h} is therefore certainly not greater than $\bar{\kappa}_h$, for which j of these conditions are given by the fixed equations (40).

We have thus the theorem: *If a quadratic form $K(x, x)$ of n variables is reduced by j linear homogeneous constraints to a quadratic form $\bar{K}(x, x)$ of $n - j$ variables, then the eigenvalues $\bar{\kappa}_1, \bar{\kappa}_2, \cdots, \bar{\kappa}_{n-j}$ of $K(x, x)$ are not greater than the corresponding numbers of the sequence $\kappa_1, \kappa_2, \cdots, \kappa_{n-j}$ and not less than the corresponding numbers of the sequence $\kappa_{j+1}, \kappa_{j+2}, \cdots, \kappa_n$.*[1]

If, in particular, we let $j = 1$ and take for our constraint the condition $x_n = 0$, then the quadratic form K goes over into its $(n - 1)$-st "*section*," and we obtain the theorem: *The h-th eigenvalue of the $(n - 1)$-st section is at most equal to the h-th eigenvalue of the original quadratic form, and at least equal to the $(h + 1)$-st eigenvalue.*

If this theorem is applied to the $(n - 1)$-st section of the quadratic form, there results a corresponding theorem for the $(n - 2)$-nd section, and so forth. In general we note that the eigenvalues of any two successive sections of a quadratic form are ordered in the indicated manner.

Moreover, we may conclude: *If a positive definite form is added*

[1] This may be illustrated geometrically: Let us consider the ellipse formed by the intersection of an ellipsoid and a plane passing through its center. The length of the major axis of this ellipse is between the lengths of the longest and the second axes of the ellipsoid, while the length of the minor axis of the ellipse is between those of the second and the shortest axes of the ellipsoid.

to $K(x, x)$, the eigenvalues of the sum are not less than the corresponding eigenvalues of $K(x, x)$.

Instead of utilizing a minimum-maximum problem to characterize the eigenvalues we may use a *maximum-minimum problem*. In this case the eigenvalues will appear in the opposite order.

It may be left to the reader to formulate and prove the minimum-maximum character of the eigenvalues of Hermitian forms.

§5. *Supplement and Problems*

1. Linear Independence and the Gram Determinant. The question of the linear dependence of m given vectors $\mathfrak{v}_1, \mathfrak{v}_2, \cdots, \mathfrak{v}_m$ may be very simply decided in the following way without explicity determining the rank of the component matrix: We consider the quadratic form

$$G(x, x) = (x_1\mathfrak{v}_1 + \cdots + x_m\mathfrak{v}_m)^2 = \sum_{i,k=1}^{m} (\mathfrak{v}_i \cdot \mathfrak{v}_k) x_i x_k .$$

Clearly $G(x, x) \geq 0$, and the vectors $\mathfrak{v}_i$ are linearly dependent if and only if there exists a set of values $x_1, x_2, \cdots, x_m$ with (25′)

$$\sum_{i=1}^{m} x_i^2 = 1,$$

for which $G(x, x) = 0$. Thus if the vectors $\mathfrak{v}_i$ are linearly dependent the minimum of the form $G(x, x)$ subject to condition (25′) must be equal to zero. But this minimum is just the smallest eigenvalue of the quadratic form $G(x, x)$, i.e. the least root of the equation

$$(41) \qquad \begin{vmatrix} \mathfrak{v}_1^2 - \kappa & (\mathfrak{v}_1 \cdot \mathfrak{v}_2) & \cdots & (\mathfrak{v}_1 \cdot \mathfrak{v}_m) \\ (\mathfrak{v}_2 \cdot \mathfrak{v}_1) & \mathfrak{v}_2^2 - \kappa & \cdots & (\mathfrak{v}_2 \cdot \mathfrak{v}_m) \\ \cdots & \cdots & \cdots & \cdots \\ (\mathfrak{v}_m \cdot \mathfrak{v}_1) & (\mathfrak{v}_m \cdot \mathfrak{v}_2) & \cdots & \mathfrak{v}_m^2 - \kappa \end{vmatrix} = 0.$$

The theorem follows:

A necessary and sufficient condition for the linear dependence of the vectors $\mathfrak{v}_1, \mathfrak{v}_2, \cdots, \mathfrak{v}_m$ is the vanishing of the "Gram determinant"

$$(42) \qquad \Gamma = \begin{vmatrix} \mathfrak{v}_1^2 & (\mathfrak{v}_1 \cdot \mathfrak{v}_2) & \cdots & (\mathfrak{v}_1 \cdot \mathfrak{v}_m) \\ (\mathfrak{v}_2 \cdot \mathfrak{v}_1) & \mathfrak{v}_2^2 & \cdots & (\mathfrak{v}_2 \cdot \mathfrak{v}_m) \\ \cdots & \cdots & \cdots & \cdots \\ (\mathfrak{v}_m \cdot \mathfrak{v}_1) & (\mathfrak{v}_m \cdot \mathfrak{v}_2) & \cdots & \mathfrak{v}_m^2 \end{vmatrix}.$$

An alternate expression for Γ follows from (41). If the left side of equation (41), which is satisfied by the (all non-negative) eigenvalues $\kappa_1, \kappa_2, \cdots, \kappa_m$ of $G(x, x)$, is developed in powers of κ, then the term independent of κ is equal to Γ, while the coefficient of κ^m is equal to $(-1)^m$. According to a well-known theorem of algebra it follows that

$$\Gamma = \kappa_1 \kappa_2 \cdots \kappa_m . \tag{43}$$

Consequently *the Gram determinant of an arbitrary system of vectors is never negative.* Relation

$$\Gamma = |(\mathbf{v}_i \cdot \mathbf{v}_k)| \geq 0 \qquad (i, k = 1, \cdots, m), \tag{44}$$

in which the equality holds only for linearly dependent vectors $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_m$, is a generalization of the Schwarz inequality (see page 2)

$$\mathbf{v}_1^2 \mathbf{v}_2^2 - (\mathbf{v}_1 \cdot \mathbf{v}_2)^2 = \begin{vmatrix} \mathbf{v}_1^2 & (\mathbf{v}_1 \cdot \mathbf{v}_2) \\ (\mathbf{v}_2 \cdot \mathbf{v}_1) & \mathbf{v}_2^2 \end{vmatrix} \geq 0.$$

The value of the Gram determinant or, alternatively, the lowest eigenvalue κ_m of the form $G(x, x)$ represents a measure of the linear independence of the vectors $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_m$. The smaller this number, the "flatter" is the m-dimensional polyhedron defined by vectors $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_m$; if it is equal to zero the polyhedron collapses into one of at most $m - 1$ dimensions. In this connection the Gram determinant has a simple geometrical significance. It is equal to the square of the $m!$-fold volume of the m-dimensional polyhedron defined by the vectors $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_m$. Thus, for $m = 2$, it is the square of twice the area of the triangle formed from $\mathbf{v}_1$ and $\mathbf{v}_2$.

Gram's criterion for linear dependence must of course be equivalent to the usual one. The latter states that vectors are linearly dependent if and only if all determinants formed with m columns of the rectangular component array

$$\begin{matrix} v_{11} & v_{12} & \cdots & v_{1n} \\ v_{21} & v_{22} & \cdots & v_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ v_{m1} & v_{m2} & \cdots & v_{mn} \end{matrix}$$

are equal to zero. And indeed, according to a well-known theorem of the theory of determinants,

$$\Gamma = \sum \begin{vmatrix} v_{1s_1} & v_{1s_2} & \cdots & v_{1s_m} \\ v_{2s_1} & v_{2s_2} & \cdots & v_{2s_m} \\ \cdots & \cdots & \cdots & \cdots \\ v_{ms_1} & v_{ms_2} & \cdots & v_{ms_m} \end{vmatrix}^2, \tag{45}$$

where the summation is extended over all integers $s_1, s_2, \cdots, s_m$ from 1 to n with $s_1 < s_2 < \cdots < s_m$.

2. Hadamard's Inequality for Determinants. Every determinant

$$\mathsf{A} = |a_{ik}| = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{2n} & \cdots & a_{nn} \end{vmatrix}$$

with real elements a_{ik} satisfies the inequality

$$\mathsf{A}^2 \leq \prod_{i=1}^{n} \sum_{k=1}^{n} a_{ik}^2. \tag{46}$$

Proof: Let the elements a_{ik} vary, keeping the sums of squares

$$\sum_{k=1}^{n} a_{ik}^2 = c_i^2 \qquad (i = 1, \cdots, n)$$

fixed. If $\mathsf{A}_{\max}^2$ is the greatest value of the function A^2 of the elements a_{ik} under these n conditions—the existence of such a maximum follows immediately from Weierstrass's theorem (see page 23)—then the elements of $\mathsf{A}_{\max}$ in each row must be proportional to the corresponding cofactors. For, if h is fixed, we have

$$\mathsf{A} = a_{h1}\mathsf{A}_{h1} + \cdots + a_{hn}\mathsf{A}_{hn};$$

thus, by the Schwarz inequality,

$$\mathsf{A}^2 \leq \sum_{k=1}^{n} a_{hk}^2 \sum_{k=1}^{n} \mathsf{A}_{hk}^2 = c_h^2 \sum_{k=1}^{n} \mathsf{A}_{hk}^2.$$

If the a_{hk} are not proportional to the A_{hk} the inequality holds, and A^2 certainly can not have its maximum value. For, in this case, by

suitably changing the n quantities a_{hk} ($k = 1, \cdots, n$), with c_h^2 and the A_{hk} held constant, the square of the determinant can be made equal to the right-hand side.

If we now multiply $\mathsf{A}_{\max}$ by itself, we obtain, according to the multiplication theorem for determinants,

$$\mathsf{A}_{\max}^2 = \prod_{i=1}^{n} c_i^2,$$

since the inner products of different rows of $\mathsf{A}_{\max}$ vanish as a result of the proportionality just demonstrated and of elementary theorems on determinants. Therefore the original determinant satisfies Hadamard's inequality

$$\mathsf{A}^2 \leq \prod_{i=1}^{n} c_i^2 = \prod_{i=1}^{n} \sum_{k=1}^{n} a_{ik}^2 .$$

The geometrical meaning of Hadamard's inequality is that the volume of the polyhedron formed from n vectors of given lengths in n-dimensional space is greatest if the vectors are mutually orthogonal.

Hadamard's inequality is also valid for complex a_{ik} if A and a_{ik} are replaced by their absolute values.

3. Generalized Treatment of Canonical Transformations. For generalizations and applications to many problems of analysis the following concise treatment of the simultaneous canonical transformation of two quadratic forms is most appropriate. Again we consider two quadratic forms in an n-dimensional vector space of vectors $\mathbf{x}, \mathbf{y}, \cdots$:

(a) $$H(x, x) = \sum_{p,q=1}^{n} h_{pq} x_p x_q ,$$

which we assume positive definite, and

(b) $$K(x, x) = \sum_{p,q=1}^{n} k_{pq} x_p x_q ,$$

which is not necessarily definite. By definition we interpret $H(x, x)$ as the square of the length of the vector $\mathbf{x}$, and the polar form

$$H(x, y) = (\mathbf{x}, \mathbf{y}) = \sum_{p,q=1}^{n} h_{pq} x_p y_q$$

as the inner product of $\mathbf{x}$ and $\mathbf{y}$. The problem is to find a linear transformation

$$x_p = \sum_{q=1}^{n} l_{pq} y_q \qquad (p = 1, \cdots, n)$$

which transforms K and H into the sums

$$K(x, x) = \sum_{p=1}^{n} \rho_p y_p^2, \qquad H(x, x) = \sum_{p=1}^{n} y_p^2.$$

To obtain this transformation explicit expressions for the forms K and H are not required; our proof is based merely on the properties that H and K are continuous functions of the vector $\mathbf{x}$, that with arbitrary constants λ and μ equations of the form

$$(47) \quad H(\lambda x + \mu y, \lambda x + \mu y) = \lambda^2 H(x, x) + 2\lambda\mu H(x, y) + \mu^2 H(y, y),$$

$$(48) \quad K(\lambda x + \mu y, \lambda x + \mu y) = \lambda^2 K(x, x) + 2\lambda\mu K(x, y) + \mu^2 K(y, y)$$

hold, and that H is positive definite, vanishing only for $\mathbf{x} = 0$.

We consider a sequence of maximum problems: First we define a vector $\mathbf{x} = \mathbf{x}^1$ for which the quotient

$$K(x, x)/H(x, x)$$

attains its maximum value ρ_1. Without affecting the value of this quotient the vector $\mathbf{x}$ may be normalized, i.e. subjected to the condition $H(x, x) = 1$.

Then we define another normalized vector $\mathbf{x}^2$ for which the quotient $K(x, x)/H(x, x)$ attains its maximum value ρ_2 under the orthogonality condition $H(x, x^1) = 0$. Proceeding in this way, we define a sequence of normalized vectors $\mathbf{x}^1, \mathbf{x}^2, \cdots, \mathbf{x}^k$, such that for $\mathbf{x} = \mathbf{x}^k$ the quotient $K(x, x)/H(x, x)$ attains its maximum value ρ_k under the orthogonality conditions

$$H(x, x^\nu) = 0 \qquad (\nu = 1, \cdots, k - 1).$$

After n steps we obtain a complete system of vectors $\mathbf{x}^1, \mathbf{x}^2, \cdots, \mathbf{x}^n$ for which the relations

$$(49) \quad H(x^i, x^k) = 1, \quad i = k; \qquad H(x^i, x^k) = 0, \quad i < k;$$

and

$$(50) \quad K(x^i, x^k) = \rho_k, \quad i = k; \qquad K(x^i, x^k) = 0, \quad i < k$$

hold. Relations (49) are merely the orthogonality relations stipulated in our maximum problems. To prove relations (50) we consider first $\mathbf{x}^1$. The maximum property of $\mathbf{x}^1$ is expressed by the inequality

$$K(x^1 + \epsilon\zeta, x^1 + \epsilon\zeta) - \rho_1 H(x^1 + \epsilon\zeta, x^1 + \epsilon\zeta) \leq 0$$

valid for an arbitrary constant ϵ and an arbitrary vector ζ. Because of (47) and (48), it yields

$$2\epsilon A + \epsilon^2 B \leq 0$$

where

$$A = K(x^1, \zeta) - \rho_1 H(x^1, \zeta), \qquad B = K(\zeta, \zeta) - \rho_1 H(\zeta, \zeta).$$

Since this inequality is valid for arbitrarily small positive or negative ϵ it implies that $A = 0$ or that

$$K(x^1, \zeta) - \rho_1 H(x^1, \zeta) = 0 \tag{51}$$

for arbitrary ζ. The maximum problem for $\mathbf{x}^h$ yields as above

$$K(x^h, \zeta) - \rho_h H(x^h, \zeta) = 0$$

for an arbitrary vector ζ satisfying the relations

$$H(\zeta, x^\nu) = 0 \qquad (\nu = 1, \cdots, h - 1).$$

Now, for $h < k$, we may take $\zeta = \mathbf{x}^k$. Since $H(x^h, x^k) = 0$, we may conclude that $K(x^h, x^k) = 0$ for $h < k$, while by definition $K(x^h, x^h) = \rho_h$.

Since the n orthogonal vectors $\mathbf{x}^\nu$ form a complete system in our vector space, an arbitrary vector $\mathbf{x}$ can be expressed in the form

$$\mathbf{x} = \sum_{\nu=1}^{n} y_\nu \mathbf{x}^\nu$$

where $y_\nu = H(x, x^\nu)$. We substitute these expressions in H and K and use the expansions corresponding to (47), (48) for n summands; because of (49), (50) it follows immediately that

$$H(x, x) = \sum_{\nu=1}^{n} y_\nu^2,$$

$$K(x, x) = \sum_{\nu=1}^{n} \rho_\nu y_\nu^2.$$

Thus we have accomplished the required transformation.

Exactly as before the values ρ_h are shown to have the following minimum-maximum property.

Under the auxiliary conditions

$$\sum_{p=1}^{n} \alpha_{\nu p} x_p = 0 \qquad (\nu = 1, \cdots, h-1),$$

ρ_h *(with* $\rho_1 \geq \cdots \geq \rho_n$*) is the least value which the maximum of* $K(x, x)/H(x, x)$ *can assume—this maximum is regarded as a function of the parameters* $\alpha_{\nu p}$.

To construct the transformation of which we have proved the existence we first show that for all integers h the "variational equation"

$$K(x^h, \zeta) - \rho_h H(x^h, \zeta) = 0$$

holds with an arbitrary vector ζ. So far the relation has been proved only under the restriction $(\zeta, \mathfrak{x}^\nu) = 0$ for $\nu < h$. However, if ζ is arbitrary the vector $\mathfrak{n} = \zeta - c_1\mathfrak{x}^1 - \cdots - c_{h-1}\mathfrak{x}^{h-1}$ with $c_\nu = (\zeta, \mathfrak{x}^\nu)$ satisfies the orthogonality condition $H(\eta, x^\nu) = 0$, $\nu < h$, hence $0 = K(x^h, \eta) - \rho_h H(x^h, \eta) = K(x^h, \zeta) - \rho_h H(x^h, \zeta)$; here the final equality sign follows from (49) and (50).

Writing the variational equation for $\mathfrak{x}^h = \mathfrak{x}$, $\rho_h = \rho$ we obtain for the components x_j of $\mathfrak{x} = \mathfrak{x}^h$ the system of linear homogeneous equations

$$\sum_{j=1}^{n} (k_{ij} - \rho h_{ij}) x_j = 0 \qquad (i = 1, \cdots, n);$$

hence the values ρ_h satisfy the determinant equation $\| k_{ij} - \rho h_{ij} \| = 0$ and the vectors $\mathfrak{x}^h$ are obtained from the linear equations after the quantities $\rho = \rho_h$ have been found. Clearly, these considerations characterize the numbers ρ_h and the vectors $\mathfrak{x}^h$ as the eigenvalues and eigenvectors of the matrix $(\mathfrak{k}_{pq})$ with respect to the matrix $(\mathfrak{h}_{pq})$.

Thus for each eigenvalue ρ_h there exists a solution in the form of a vector $\mathfrak{x}^h$. The solutions for different eigenvalues are orthogonal; if two eigenvalues are equal the corresponding solutions are not necessarily orthogonal but may be made so by the orthogonalization process of page 4. These mutually orthogonal solutions may be normalized to unit length; the resulting vectors are the eigenvectors of the problem and their components are the coefficients of the required transformation.

These coefficients l_{pq} are obtained from $\mathbf{x} = \sum_{q=1}^{n} y_q \mathbf{x}^q$ if we multiply by the vector $\mathbf{e}^p$ which defines the original coordinate system. Thus $x_p = (\mathbf{x}, \mathbf{e}^p) = \sum_{q=1}^{n} y_q(\mathbf{x}^q, \mathbf{e}^p)$; hence $l_{pq} = (\mathbf{x}^q, \mathbf{e}^p)$.

4. Bilinear and Quadratic Forms of Infinitely Many Variables. Under suitable conditions our theory remains valid if the number of variables increases beyond all bounds. For example, this is the case if both the sum of the squares of the coefficients of the bilinear or quadratic forms and the sum of the squares of the variables converge. This theory of forms of infinitely many variables, developed by Hilbert, may then be applied to numerous problems of analysis. However, the theory of forms in vector spaces of infinitely many dimensions can be more adequately developed on the basis of abstract concepts as indicated in subsection 3. As we shall see, many topics in analysis can be illuminated from the viewpoint of such a generalized theory of quadratic forms.

5. Infinitesimal Linear Transformations. An infinitesimal linear transformation is defined as a transformation whose matrix is

$$A = E + (\epsilon\alpha_{ik}) = \begin{pmatrix} 1 + \epsilon\alpha_{11} & \epsilon\alpha_{12} & \cdots & \epsilon\alpha_{1n} \\ \epsilon\alpha_{21} & 1 + \epsilon\alpha_{22} & \cdots & \epsilon\alpha_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ \epsilon\alpha_{n1} & \epsilon\alpha_{n2} & \cdots & 1 + \epsilon\alpha_{nn} \end{pmatrix},$$

where ϵ denotes an infinitesimal quantity of the first order, i.e. a quantity whose higher powers are, for the problem at hand, negligible in comparison with lower powers of ϵ. The product of two such infinitesimal transformations with the matrices $A = E + (\epsilon\alpha_{ik})$ and $B = E + (\epsilon\beta_{ik})$ has the matrix $C = E + (\epsilon\alpha_{ik} + \epsilon\beta_{ik})$. Thus the product does not depend on the order of the factors; in other words, *infinitesimal transformations commute with each other.*

Furthermore, *the reciprocal matrix of* $A = E + (\epsilon\alpha_{ik})$ *is* $A^{-1} = E - (\epsilon\alpha_{ik})$, *and the determinant of the matrix A is equal to*

$$1 + \epsilon(\alpha_{11} + \alpha_{22} + \cdots + \alpha_{nn}).$$

If the infinitesimal transformation is to be orthogonal, we have the condition $A'A = E$, where A' is the transposed matrix. We must therefore have $\alpha_{ik} + \alpha_{ki} = 0$, or, in other words:

A necessary and sufficient condition for the orthogonality of an in-

finitesimal transformation is that the difference between its matrix and the unit matrix be skew-symmetric.

Any infinitesimal transformation with the matrix $C = E + (\epsilon\gamma_{ik})$ may be represented as the product of an orthogonal transformation $A = E + (\epsilon\alpha_{ik})$ and a symmetric transformation $B = E + (\epsilon\beta_{ik})$, where

$$\alpha_{ik} = \tfrac{1}{2}(\gamma_{ik} - \gamma_{ki}),$$

$$\beta_{ik} = \tfrac{1}{2}(\gamma_{ik} + \gamma_{ki}).$$

Consider a symmetric transformation $y_i = \sum_k s_{ik}x_k$ whose matrix is $S = (s_{ik})$, not necessarily infinitesimal. Its geometrical significance is that of a dilatation in n mutually orthogonal directions. To see this let us transform the quadratic form $S(x, x)$ to principal axes, transforming the x_i into u_i and the y_i into v_i. We then have

$$\sum_{i,k=1}^{n} s_{ik}x_ix_k = \sum_{i=1}^{n} \kappa_i u_i^2,$$

and the equations $y_i = \sum_k s_{ik}x_k$ become

$$v_i = \kappa_i u_i .$$

These equations evidently represent a dilatation by the factor κ_i in the direction of the i-th principal axis. The ratio of the increase of volume to the initial volume, known as the volume dilatation, is evidently given by the difference $\kappa_1\kappa_2 \cdots \kappa_n - 1 = |\, s_{ik} \,| - 1$. If, in particular, the transformation is infinitesimal, i.e. $(s_{ik}) = E + (\epsilon\beta_{ik})$, we have

$$\kappa_1 \cdots \kappa_n - 1 = \epsilon(\beta_{11} + \cdots + \beta_{nn}).$$

Since an orthogonal transformation represents a rotation we may summarize by stating:

An infinitesimal transformation whose matrix is $E + (\epsilon\gamma_{ik})$ may be represented as the product of a rotation and a dilatation; the volume dilatation is $\epsilon \sum_{i=1}^{n} \gamma_{ii}$.

6. Perturbations. In the theory of small vibrations and in many problems of quantum mechanics it is important to determine how the eigenvalues and eigenvectors of a quadratic form $K(x, x) = \sum_{i,k=1}^{n} b_{ik}x_ix_k$ are changed if both the form $K(x, x)$ and the unit

form $E(x, x)$ are altered. Suppose $E(x, x)$ is replaced by $E(x, x) + \epsilon \mathsf{A}(x, x)$ and $K(x, x)$ by $K(x, x) + \epsilon \mathsf{B}(x, x)$, where

$$\mathsf{A}(x, x) = \sum_{i,k=1}^{n} \alpha_{ik} x_i x_k, \qquad \mathsf{B}(x, x) = \sum_{i,k=1}^{n} \beta_{ik} x_i x_k,$$

and ϵ is a parameter. The problem is then to transform $E + \epsilon\mathsf{A}$ and $K + \epsilon\mathsf{B}$ simultaneously into canonical form. If we put

$$K(x, x) + \epsilon \mathsf{B}(x, x) = \sum_{i,k=1}^{n} b'_{ik} x_i x_k,$$

$$E(x, x) + \epsilon \mathsf{A}(x, x) = \sum_{i,k=1}^{n} a'_{ik} x_i x_k,$$

the equations for the components of the eigenvectors become

$$\sum_{k=1}^{n} (b'_{ik} - \rho' a'_{ik}) x'_k = 0 \qquad (i = 1, \cdots, n),$$

where ρ' may be obtained from the condition that the determinant of this system of equations must vanish. Let us denote the eigenvalues of $K(x, x)$ by $\rho_1, \rho_2, \cdots, \rho_n$ and assume that they are all different; let the corresponding values for the varied system be denoted by $\rho'_1, \rho'_2, \cdots, \rho'_n$. The original form $K(x, x)$ may be assumed to be a sum of squares:

$$K(x, x) = \rho_1 x_1^2 + \rho_2 x_2^2 + \cdots + \rho_n x_n^2.$$

The quantities ρ'_i, being simple roots of an algebraic equation, are single-valued analytic functions of ϵ in the neighborhood of $\epsilon = 0$; the same is, therefore, true of the components x'_{hk} of the varied eigenvectors belonging to the eigenvalues ρ'_h. Thus the quantities ρ'_h and x'_{hk} may be expressed as power series in ϵ, the constant terms of which are, of course, the original eigenvalues ρ_h and the components of the original eigenvectors x_{hk}, respectively. In order to compute successively the coefficients of $\epsilon, \epsilon^2, \cdots$ we must substitute these power series in the equations

$$\sum_{k=1}^{n} (b'_{ik} - \rho'_h a'_{ik}) x'_{hk} = 0 \qquad (i, h = 1, \cdots, n)$$

in which we have $b'_{ik} = \rho_{ik} + \epsilon\beta_{ik}$, $a'_{ik} = \delta_{ik} + \epsilon\alpha_{ik}$, with $\rho_{ii} = \rho_i$, $\rho_{ik} = 0$ $(i \neq k)$, $\delta_{ii} = 1$, $\delta_{ik} = 0$ $(i \neq k)$. By collecting the terms

in each power of ϵ in these equations and then setting the coefficient of each power of ϵ equal to zero we obtain an infinite sequence of new equations. An equivalent procedure, which is often somewhat more convenient, is given by the following considerations of orders of magnitude, in which ϵ is regarded as an infinitesimal quantity. We first consider the equation with $i = h$. By setting the coefficient of the first power of ϵ equal to zero we obtain

$$\rho_h' = \frac{\rho_h + \epsilon\beta_{hh}}{1 + \epsilon\alpha_{hh}} = \rho_h - \epsilon\rho_h\alpha_{hh} + \epsilon\beta_{hh},$$

except for terms of the second or higher orders in ϵ. The same procedure applied to the equations with $i \neq h$ yields the result

$$x_{hh}' = 1, \qquad x_{hi}' = -\epsilon\,\frac{\alpha_{ih}\rho_h - \beta_{ih}}{\rho_h - \rho_i},$$

except for infinitesimal quantities of the second order in ϵ.

By using these values of the components of the eigenvectors we may easily obtain the eigenvalues up to and including the second order in ϵ. Again we consider the h-th equation for the components of the h-th eigenvector:

$$\sum_{k=1}^{n} (b_{hk}' - \rho_h' a_{hk}')x_{hk}' = 0.$$

If we neglect quantities of the third order in ϵ on the left-hand side and write the term with $h = k$ separately, we obtain

$$\begin{aligned} b_{hh}' - \rho_h' a_{hh}' &= \sum_{k=1}^{n}{}' \epsilon(b_{hk}' - \rho_h' a_{hk}')\,\frac{\alpha_{kh}\rho_h - \beta_{kh}}{\rho_h - \rho_k} \\ &= -\epsilon^2 \sum_{k=1}^{n}{}' \frac{(\alpha_{kh}\rho_h - \beta_{kh})^2}{\rho_h - \rho_k}. \end{aligned}$$

It follows that

$$\rho_h' = \rho_h - \epsilon(\rho_h\alpha_{hh} - \beta_{hh}) - \epsilon^2\alpha_{hh}(\beta_{hh} - \rho_h\alpha_{hh}) + \epsilon^2 \sum_{k=1}^{n}{}' \frac{(\alpha_{kh}\rho_h - \beta_{kh})^2}{\rho_h - \rho_k}.$$

Here we have used the symbol $\sum_k'$ to denote summation over all values of k from 1 to n except for $k = h$.

7. Constraints. Constraints expressed by linear conditions

$$\gamma_1 x_1 + \cdots + \gamma_n x_n = 0,$$

and the resulting diminution of the number of independent variables of the quadratic form $K(x, x) = \sum_{p,q=1}^{n} k_{pq}x_p x_q$, may be regarded as the end result of a continuous process. Consider the quadratic form $K(x, x) + t(\gamma_1 x_1 + \cdots + \gamma_n x_n)^2$, where t is a positive parameter. If t increases beyond all bounds, each eigenvalue increases monotonically. The greatest eigenvalue increases beyond all bounds, while the others approach the eigenvalues of the quadratic form which is obtained from $K(x, x)$ by elimination of one variable in accordance with the given constraint.

8. Elementary Divisors of a Matrix or a Bilinear Form. Let $\mathbf{A}$ be a tensor and $A = (a_{ik})$ the corresponding matrix. Then the polynomial

$$|\kappa E - A| = \begin{vmatrix} \kappa - a_{11} & -a_{12} & \cdots & -a_{1n} \\ -a_{21} & \kappa - a_{22} & \cdots & -a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ -a_{n1} & -a_{n2} & \cdots & \kappa - a_{nn} \end{vmatrix}$$

may be decomposed according to certain well-known rules into the product of its "elementary divisors"

$$(\kappa - r_1)^{e_1}, (\kappa - r_2)^{e_2}, \cdots, (\kappa - r_h)^{e_h},$$

where some of the numbers $r_1, r_2, \cdots, r_h$ may be equal. For each divisor $(\kappa - r_\nu)^{e_\nu}$ there is a system of e_ν vectors $\mathbf{f}_1^{(\nu)}, \mathbf{f}_2^{(\nu)}, \cdots, \mathbf{f}_{e_\nu}^{(\nu)}$ such that the equations

$$\mathbf{A}\mathbf{f}_1^{(\nu)} = r_\nu \mathbf{f}_1^{(\nu)}, \quad \mathbf{A}\mathbf{f}_2^{(\nu)} = r_\nu \mathbf{f}_2^{(\nu)} + \mathbf{f}_1^{(\nu)}, \cdots, \mathbf{A}\mathbf{f}_{e_\nu}^{(\nu)} = r_\nu \mathbf{f}_{e_\nu}^{(\nu)} + \mathbf{f}_{e_\nu - 1}^{(\nu)}$$

are valid. Here the n vectors

$$\mathbf{f}_1^{(1)}, \cdots, \mathbf{f}_{e_1}^{(1)}; \quad \mathbf{f}_1^{(2)}, \cdots, \mathbf{f}_{e_2}^{(2)}; \cdots; \mathbf{f}_1^{(h)}, \cdots, \mathbf{f}_{e_h}^{(h)}$$

are linearly independent. If they are introduced as new variables $x_1^{(1)}, x_2^{(2)}, \cdots, x_{e_h}^{(h)}$, the matrix A is transformed into the matrix

$$\begin{pmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & A_h \end{pmatrix}$$

in which $A_1, A_2, \cdots, A_h$ are themselves matrices; A_ν is a matrix of order e_ν :

$$A_\nu = \begin{pmatrix} r_\nu & 0 & 0 & \cdots & 0 & 0 \\ 1 & r_\nu & 0 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 1 & r_\nu \end{pmatrix}.$$

9. Spectrum of a Unitary Matrix. We shall now show that the spectrum of a unitary matrix lies on the unit circle, i.e. that all of its eigenvalues have the absolute value 1.

We note that the elements of a unitary matrix cannot exceed unity in absolute value. Therefore the absolute values of the coefficients of the characteristic equations of all unitary matrices of the n-th degree must lie below a certain bound which is independent of the particular matrix considered. Since the absolute values of the first and last coefficients of the characteristic equation are equal to 1, this means that the absolute values of the eigenvalues must lie between certain positive upper and lower bounds which are independent of the particular matrix. On the other hand all powers A^m of a unitary matrix A are also unitary, and their eigenvalues are the m-th powers of the corresponding eigenvalues of A. But the absolute values of these powers and their reciprocals can remain below a bound which is independent of m only if the absolute value of each eigenvalue (and all of its powers) is 1.

Another proof, which can be used for infinite matrices as well, follows from the convergence of the Neumann series for $(E - \lambda A)^{-1}$. The series

$$(E - \lambda A)^{-1} = E + \lambda A + \lambda^2 A^2 + \cdots,$$

where A is a unitary matrix, certainly converges if $|\lambda| < 1$. For the elements of the matrices A^m all have absolute values of at most 1, and thus the geometric series is a dominating series for the matrix elements. Thus no zeros of $|E - \lambda A|$ can lie inside the unit circle On the other hand we have, in virtue of the relation $A\bar{A}' = E$,

$$(E - \lambda A)^{-1} = -\frac{1}{\lambda}\overline{A'}\left(E + \frac{1}{\lambda}\overline{A}' + \frac{1}{\lambda^2}\overline{A'^2} + \cdots\right).$$

Here the geometric series on the right converges for $| 1/\lambda | < 1$ since A' is also a unitary matrix. Thus no zero of $| E - \lambda A |$ can lie outside the unit circle. Therefore all these zeros lie on the unit circle, and our assertion is proved.

References

Textbooks

Bôcher, M., Introduction to Higher Algebra. Macmillan, New York, 1907.

Kowalewski, G., Einführung in die Determinantentheorie. Veit, Leipzig, 1909.

Wintner, A., Spektraltheorie der unendlichen Matrizen. S. Hirzel, Leipzig, 1929.

Monographs and Articles

Courant, R., Zur Theorie der kleinen Schwingungen. Zts. f. angew. Math. u. Mech., Vol. 2, 1922, pp. 278–285.

Fischer, E., Über quadratische Formen mit reelen Koeffizienten. Monatsh. f. Math. u. Phys., Vol. 16, 1905, pp. 234–249. The maximum-minimum character of the eigenvalues probably was first mentioned in this paper.

Hilbert, D., Grundzüge einer allgemeinen Theorie der linearen Integralgleichungen, especially sections 1 and 4. Fortschritte der mathematischen Wissenschaften in Monographien, Heft 3. B. G. Teubner, Leipzig and Berlin, 1912.

CHAPTER II

Series Expansions of Arbitrary Functions

Many of the relations discussed in Chapter I are closely analogous to theorems in which functions of one or more variables, defined in some given *fundamental domain* G, take the place of vectors in n-dimensional space. The problem of expressing a vector in n-dimensional space as a linear combination of n independent but otherwise arbitrary vectors is analogous to the problem of representing a more or less arbitrary function in the domain G as a linear combination of members of some given set of functions. (It will become obvious that the number of functions in this set must be infinite.) This is known as the *problem of series expansion of an arbitrary function in terms of a given set of functions*. In the present chapter this problem, which appears in various forms throughout mathematical physics, will be treated from a general point of view.

We shall restrict ourselves to *piecewise continuous functions*; i.e. we consider functions for which the fundamental domain G may be so subdivided into a finite number of domains that, in the interior of each domain, the function is continuous and approaches a finite limit as a point on the boundary of one of these domains is approached from its interior. To simplify notation we shall at first consider functions of one variable x, whose fundamental domain G is a finite interval on the x-axis. If we are concerned with several variables, say the two variables x and y, we shall assume that the boundary of the fundamental domain G is formed by a finite number of segments of curves with continuously turning tangents. If we consider the points on the boundary as belonging to the fundamental domain we shall speak of a "closed domain."

Furthermore we shall in many cases assume that the functions considered are *piecewise smooth*, i.e. that they are piecewise continuous and possess piecewise continuous first derivatives. Unless the contrary is specified explicitly we shall assume that our functions have real values.

§1. *Orthogonal Systems of Functions*

1. Definitions. The integral[1]

$$(f, g) = \int fg\, dx \tag{1}$$

taken over the (finite) fundamental domain is called the *inner product* (f, g) *or* (fg) *of two functions* $f(x)$ *and* $g(x)$. It satisfies the *Schwarz inequality*

$$(f, g)^2 \leq (f, f)(g, g), \tag{2}$$

where the equality holds if and only if f and g are proportional. As in the case of vectors, this follows either from the positive definite character of the function $\int (\lambda f + g)^2\, dx$ of the variable λ, or directly from the identity

$$(f, g)^2 = (f, f)(g, g) - \tfrac{1}{2} \iint (f(x)g(\xi) - f(\xi)g(x))^2\, dx\, d\xi.$$

Two functions $f(x)$ and $g(x)$ for which (f, g) vanishes will be termed *orthogonal*. The inner product of a function $f(x)$ with itself is called the *norm* of this function and is denoted by Nf:

$$Nf = (f, f) = \int f^2\, dx. \tag{3}$$

A function whose norm is unity is said to be *normalized*. A system of normalized functions $\varphi_1(x), \varphi_2(x), \cdots$ any two different members of which are orthogonal is called an *orthonormal system* and the relations

$$(\varphi_\nu, \varphi_\mu) = \delta_{\nu\mu} \qquad (\delta_{\nu\nu} = 1, \quad \delta_{\nu\mu} = 0 \text{ for } \nu \neq \mu)$$

expressing this property are the *orthogonality relations*.

An example of an orthonormal system of functions in the interval $0 \leq x \leq 2\pi$, or more generally in any interval of length 2π, is given by the functions

$$\frac{1}{\sqrt{2\pi}}, \quad \frac{\cos x}{\sqrt{\pi}}, \quad \frac{\sin x}{\sqrt{\pi}}, \quad \frac{\cos 2x}{\sqrt{\pi}}, \quad \frac{\sin 2x}{\sqrt{\pi}}, \quad \cdots.$$

For functions of a real variable which take on *complex* values it is

[1] If no misunderstanding is possible, limits of integration will be omitted

convenient to extend the concept of orthogonality in the following way: Two complex functions $f(x)$ and $g(x)$ are said to be orthogonal if the relations

$$(f, \bar{g}) = (\bar{f}, g) = 0$$

hold, where $\bar{f}$ and $\bar{g}$ denote the complex conjugate functions to f and g, respectively. The function $f(x)$ is said to be normalized if $Nf = \int |f|^2 \, dx = 1$. The simplest example of a complex orthonormal system is given by the exponential functions

$$\frac{1}{\sqrt{2\pi}}, \quad \frac{e^{ix}}{\sqrt{2\pi}}, \quad \frac{e^{2ix}}{\sqrt{2\pi}}, \quad \cdots$$

in the interval $0 \leq x \leq 2\pi$, as is seen immediately from the "*orthogonality relations*"

$$(4) \qquad \frac{1}{2\pi} \int_0^{2\pi} e^{i(\mu-\nu)x} \, dx = \delta_{\mu\nu} \qquad (\delta_{\nu\nu} = 1, \delta_{\mu\nu} = 0 \text{ for } \mu \neq \nu).$$

We say that r functions $f_1, f_2, \cdots, f_r$ are *linearly dependent* if a homogeneous linear relation $\sum_{i=1}^{r} c_i f_i = 0$ with constant coefficients c_i, not all equal to zero, holds for all x. Otherwise the functions are said to be *linearly independent*. It is worth noting that the functions of an orthogonal system are always linearly independent. For, if an identity

$$c_1\varphi_1 + c_2\varphi_2 + \cdots + c_n\varphi_n = 0$$

holds, we may multiply by φ_ν and integrate, obtaining the result $c_\nu = 0$.

2. Orthogonalization of Functions. From a given system of infinitely many functions $v_1, v_2, \cdots$, any r of which are linearly independent for arbitrary r, an orthonormal system $\varphi_1, \varphi_2, \cdots$ may be obtained by taking φ_n as a suitable linear combination of $v_1, v_2, \cdots, v_n$. The orthogonalization procedure is exactly analogous to the procedure used for obtaining an orthogonal system of vectors from a set of linearly independent vectors (cf. Chapter I). We start by choosing $\varphi_1 = v_1/\sqrt{Nv_1}$. Next we find a number c_1' in such a way that the function $v_2 - c_1'\varphi_1$ is orthogonal to φ_1, i.e. we set $c_1' = (\varphi_1 v_2)$; $v_2 - c_1'\varphi_1$ cannot vanish identically because of the linear independence of v_1 and v_2 and, therefore, of φ_1 and v_2. We may then divide

this function by the square root of its norm, obtaining a normalized function φ_2 which is orthogonal to φ_1. We next find two numbers c_1'', c_2'' in such a way that the function $v_3 - c_1''\varphi_1 - c_2''\varphi_2$ is orthogonal to both φ_1 and φ_2, i.e. we set $c_1'' = (\varphi_1 v_3)$ and $c_2'' = (\varphi_2 v_3)$. This function cannot vanish identically and can, therefore, be normalized; we divide it by the square root of its norm and obtain the normalized function φ_3. Continuing this procedure indefinitely we arrive at the desired orthogonal system of functions.

In speaking of orthogonalization, we shall, in general, mean the procedure just described in which functions are both normalized and orthogonalized.

3. Bessel's Inequality. Completeness Relation. Approximation in the Mean. Let $\varphi_1, \varphi_2, \cdots$ be an orthonormal system and let f be any function. The numbers

$$c_\nu = (f\varphi_\nu) \qquad (\nu = 1, 2, \cdots) \tag{5}$$

are called the *expansion coefficients* or *components* of f with respect to the given orthonormal system.[1]

From the obvious relation

$$\int \left(f - \sum_{\nu=1}^{n} c_\nu \varphi_\nu\right)^2 dx \geq 0 \tag{6}$$

we obtain, by writing out the square and integrating term by term,

$$0 \leq \int f^2\, dx - 2\sum_{\nu=1}^{n} c_\nu \int f\varphi_\nu\, dx + \sum_{\nu=1}^{n} c_\nu^2 = Nf - 2\sum_{\nu=1}^{n} c_\nu^2 + \sum_{\nu=1}^{n} c_\nu^2,$$

and hence

$$\sum_{\nu=1}^{n} c_\nu^2 \leq Nf. \tag{7}$$

Since the number on the right is independent of n, it follows that

$$\sum_{\nu=1}^{\infty} c_\nu^2 \leq Nf. \tag{8}$$

This fundamental inequality, known as "*Bessel's inequality*," is true for every orthonormal system. It proves that the sum of the squares of the expansion coefficients always converges.

[1] The term "Fourier coefficients" is also used since the expansion considered is a generalization of the Fourier series expansion.

For systems of functions with complex values the corresponding relation

$$\sum_{\nu=1}^{\infty} |\, c_\nu \,|^2 \leq Nf = (f, \bar{f}) \tag{8'}$$

holds, where c_ν is the expansion coefficient $c_\nu = (f, \bar{\varphi}_\nu)$. This relation may be obtained from the inequality

$$\int |\, f(x) - \sum_{\nu=1}^{n} c_\nu \varphi_\nu \,|^2 \, dx = Nf - \sum_{\nu=1}^{n} |\, c_\nu \,|^2 \geq 0.$$

The significance of the integral in (6) is that it occurs in the problem of *approximating the given function* $f(x)$ *by a linear combination* $\sum_{\nu=1}^{n} \gamma_\nu \varphi_\nu$ *with constant coefficients* γ_ν *and fixed* n, *in such a way that the "mean square error"* $M = \int (f - \sum_{\nu=1}^{n} \gamma_\nu \varphi_\nu)^2 \, dx$ *is as small as possible.* For, by simple manipulations of the integral, we obtain the identity

$$M = \int \left(f - \sum_{\nu=1}^{n} \gamma_\nu \varphi_\nu \right)^2 dx = \int f^2 \, dx + \sum_{\nu=1}^{n} (\gamma_\nu - c_\nu)^2 - \sum_{\nu=1}^{n} c_\nu^2 ,$$

from which it follows immediately that M takes on its least value for $\gamma_\nu = c_\nu$.

An approximation of this type is known as an approximation by the method of least squares, or an *approximation "in the mean."*

If, for a given orthonormal system $\varphi_1, \varphi_2, \cdots$, any piecewise continuous function f can be approximated in the mean to any desired degree of accuracy by choosing n large enough, i.e. if n may be so chosen that the mean square error $\int (f - \sum_{\nu=1}^{n} c_\nu \varphi_\nu)^2 \, dx$ is less than a given arbitrarily small positive number, then the system of functions $\varphi_1, \varphi_2, \cdots$ is said to be *"complete."* For a complete orthonormal system of functions Bessel's inequality becomes an equality for every function f:

$$\sum_{\nu=1}^{\infty} c_\nu^2 = Nf. \tag{9}$$

This relation is known as the *"completeness relation."* It may be written in the more general form

$$\sum_{\nu=1}^{\infty} c_\nu d_\nu = (f, g) \text{ with } c_\nu = (f, \varphi_\nu),\ d_\nu = (g, \varphi_\nu), \tag{9'}$$

as can be seen by applying (9) to the function $f + g$:

$$N(f+g) = Nf + Ng + 2(f,g) = \sum_{\nu=1}^{\infty} (c_\nu + d_\nu)^2$$

$$= \sum_{\nu=1}^{\infty} (c_\nu^2 + d_\nu^2 + 2c_\nu d_\nu),$$

and then subtracting the corresponding equations for f and g.

Incidentally, a sufficient condition for the completeness of a system of functions $\varphi_1, \varphi_2, \cdots$ is that the completeness relation (9) be satisfied for all *continuous* functions f. For, any piecewise continuous function g may be approximated by a continuous function f in such a way that the integral $\int (f-g)^2\,dx$ is arbitrarily small. Let the points of discontinuity of g be denoted by x_ν. An approximating continuous function f is obtained if in each interval $x_\nu - \delta \leq x \leq x_\nu + \delta$ we replace the graph of g by the straight line-segment that joins the points $(x_\nu - \delta, g(x_\nu - \delta))$ and $(x_\nu + \delta, g(x_\nu + \delta))$.[1]

If n is sufficiently large, the mean square error integral $\int (f - \sum_{\nu=1}^{n} c_\nu\varphi_\nu)^2\,dx$, where $c_1, c_2, \cdots$ are the expansion coefficients of f, will be arbitrarily small. This implies that the integral

$$M' = \int \left(g - \sum_{\nu=1}^{n} c_\nu\varphi_\nu\right)^2 dx = \int \left[(g-f) + \left(f - \sum_{\nu=1}^{n} c_\nu\varphi_\nu\right)\right]^2 dx$$

may be made arbitrarily close to $N(g-f)$ by a suitable choice of n. From the Schwarz inequality we have, in fact,

$$M' = N(g-f) + N\left(f - \sum_{\nu=1}^{n} c_\nu\varphi_\nu\right) + 2\left(g-f, f - \sum_{\nu=1}^{n} c_\nu\varphi_\nu\right)$$

$$\leq N(g-f) + N\left(f - \sum_{\nu=1}^{n} c_\nu\varphi_\nu\right) + 2\sqrt{N(g-f)\cdot N\left(f - \sum c_\nu\varphi_\nu\right)}.$$

Moreover,

$$M = \int \left(g - \sum_{\nu=1}^{n} a_\nu\varphi_\nu\right)^2 dx \leq M',$$

[1] If M is an upper bound of $|g(x)|$ and q the number of discontinuities of g in the interval of integration, the mean square error is

$$\int (f-g)^2\,dx \leq 8M^2q\delta,$$

which can be made arbitrarily small by choosing δ small enough.

where the a_ν are the expansion coefficients for g, since these coefficients yield the least mean square error for g. Thus the completeness relation holds for g if it holds for f.

It is important to bear in mind that the completeness of an orthonormal system $\varphi_1, \varphi_2, \cdots$, expressed by the equation

$$\lim_{n\to\infty} \int \left(f - \sum_{\nu=1}^{n} c_\nu \varphi_\nu\right)^2 dx = 0,$$

does not necessarily imply that $f = \sum_{\nu=1}^{\infty} c_\nu \varphi_\nu$, i.e. that f can be expanded in a series in the functions φ_ν. *The expansion is, however, valid if the series* $\sum_{\nu=1}^{\infty} c_\nu \varphi_\nu$ *converges uniformly; the passage to the limit may then be carried out under the integral sign.* The completeness of the system $\varphi_1, \varphi_2, \cdots$ is, of course, a necessary condition for the validity of the expansion in general; for example, if we take f to be one function of a complete system, then all its components with respect to the incomplete system (consisting of all the functions φ_ν except f) will vanish. But, even for a complete system $\varphi_1, \varphi_2, \cdots$, this question of the convergence of the series for an arbitrary function requires a more detailed investigation, which will be carried out subsequently (Chapters V and VI).

If the above limit equation is satisfied we say that *the functions* $\sum_{\nu=1}^{n} c_\nu \varphi_\nu$ *converge to the function* f *in the mean.*

Another important theorem is that *a piecewise continous function is uniquely determined by its expansion coefficients with respect to a given complete orthonormal system.* That is, if two piecewise continuous functions have the same expansion coefficients they are identical. For, the difference of two functions with equal coefficients has the coefficients zero; according to the completeness relation its norm is therefore zero, and this difference vanishes identically. Thus a function is uniquely determined by its expansion in terms of a complete orthonormal system even if this expansion converges only in the mean. In fact, convergence in the mean will be all that is needed for many theorems of mathematical physics.

The concept of completeness of a system of functions retains its meaning even if the system is not orthonormal. *In general, we call a system of functions complete if every piecewise continuous function can be approximated in the mean arbitrarily closely by a linear combination of functions of the system.* The completeness of such a system is preserved under the process of orthogonalization.

4. Orthogonal and Unitary Transformations with Infinitely Many Variables. There are many analogies between orthonormal systems of functions and orthonormal systems of vectors in n-dimensional space. It is often useful to regard the functions $\varphi_1, \varphi_2, \cdots$ as "coordinate vectors" or "coordinate functions" in a space of infinitely many dimensions (often called a "Hilbert space"). An arbitrary function f is a vector in this space and its expansion coefficients $c_\nu = (f\varphi_\nu)$ are the components of this vector in terms of the system of coordinates defined by $\varphi_1, \varphi_2, \cdots$.

If $\psi_1, \psi_2, \cdots$ is a second orthonormal system of functions with respect to which the components of f are $d_\nu = (f\psi_\nu)$, and if both systems are complete, then the c_i and d_i are related by the system of infinitely many equations

$$c_i = \sum_{k=1}^{\infty} a_{ik} d_k, \qquad a_{ik} = (\varphi_i \psi_k) \qquad (i = 1, 2, \cdots). \tag{10}$$

This is seen by applying the completeness relation (9′) to the expansion coefficients of the functions f and φ_i with respect to the system $\psi_1, \psi_2, \cdots$. In the same way we obtain the inverse set of equations

$$d_i = \sum_{k=1}^{\infty} a_{ki} c_k, \qquad a_{ki} = (\psi_i \varphi_k) \qquad (i = 1, 2, \cdots). \tag{10′}$$

The coefficients obey the conditions

$$\sum_{k=1}^{\infty} a_{ik} a_{jk} = (\varphi_i \varphi_j) = \delta_{ij} \tag{11}$$

$$\sum_{k=1}^{\infty} a_{ki} a_{kj} = (\psi_i \psi_j) = \delta_{ij}, \tag{11′}$$

which are simply orthogonality conditions in n-dimensional space (Ch. I, §1) generalized to the space of infinitely many dimensions. We therefore call a transformation (10) which fulfills conditions (11) and (11′) an *orthogonal transformation of infinitely many variables* or an orthogonal transformation in Hilbert space.

Analogously, the expansion coefficients of a function with respect to two different complex orthogonal systems are related by a *unitary transformation of infinitely many variables.*

5. Validity of the Results for Several Independent Variables. More General Assumptions. None of our concepts or results is changed if, instead of functions of a single variable, we consider functions of several variables, say x and y. The variables are assumed to lie within a finite domain G, the volume element of which we denote by dG. We define the inner product (fg) of two functions $f(x, y)$ and $g(x, y)$ in the region G as the integral $(fg) = \int_G fg dG$. Then essentially nothing need be changed in the notations and proofs of this section.

Furthermore, all our concepts and results remain valid if the fundamental domain is infinite, provided we assume that the squares of all functions occurring in the treatment are integrable over the entire fundamental domain.

Finally, we remark that our methods are applicable even if the function f becomes infinite in the fundamental domain, provided that f and its square are integrable over the fundamental domain.

6. Construction of Complete Systems of Functions of Several Variables. If complete systems of functions of one variable are known, it is possible to construct complete systems of functions of two or more variables by the following theorem:

Let

$$\varphi_1(s), \varphi_2(s), \cdots$$

be a complete orthonormal system of functions in the interval $a \leq s \leq b$, and suppose for each i $(i = 1, 2, \cdots)$, that

$$\psi_{1i}(t), \psi_{2i}(t), \cdots$$

is a similar system in the interval $c \leq t \leq d$. Then the functions

$$\omega_{ik}(s, t) = \varphi_i(s)\psi_{ki}(t)$$

form a complete orthonormal system of functions in s and t in the rectangle $a \leq s \leq b$, $c \leq t \leq d$. (In particular, the system of functions $\varphi_i(s)\varphi_k(t)$ is orthonormal and complete in the square $a \leq s \leq b$, $a \leq t \leq b$.) Specifically, if $f(s, t)$ is a continuous function in this rectangle, the completeness relation

$$\iint f^2(s, t)\, ds\, dt = \sum_{i,k=1}^{\infty} \left(\iint f(s, t)\omega_{ik}(s, t)\, ds\, dt \right)^2$$

is satisfied.

To prove this theorem, we note that $\int f^2(s, t)\, ds = \sum_{i=1}^{\infty} g_i^2(t)$, where $g_i(t) = \int f(s, t)\varphi_i(s)\, ds$. This is simply an expression of the completeness of the system φ_i. Since the series on the right converges uniformly,[1] we may integrate term by term with respect to t, obtaining

$$\iint f^2(s, t)\, ds\, dt = \sum_{i=1}^{\infty} \int g_i^2(t)\, dt.$$

Now to the i-th term on the right we apply the completeness relation for the system of functions $\psi_{ki}(t)$ $(k = 1, 2, \cdots)$ and immediately obtain the desired completeness relation.

§2. *The Accumulation Principle for Functions*

1. Convergence in Function Space. The analogy between functions and vectors in n-dimensional space breaks down in many ways if infinite sets of functions and vectors are considered. It follows from an elementary theorem of analysis (Weierstrass theorem on points of accumulation) that one can select a convergent sequence out of any infinite set of vectors $\mathbf{v}$ with bounded absolute values $|\mathbf{v}|$ or bounded norms $\mathbf{v}^2 = N\mathbf{v}$. Moreover, if for a sequence of vectors $\mathbf{v}_1, \mathbf{v}_2, \cdots$ the relation $\lim_{m\to\infty}^{n\to\infty} N(\mathbf{v}_n - \mathbf{v}_m) = 0$ holds, then a limit vector $\mathbf{v} = \lim_{n\to\infty} \mathbf{v}_n$ exists; finally, $\lim_{n\to\infty} N\mathbf{v}_n = 0$ implies $\lim_{n\to\infty} \mathbf{v}_n = 0$. These statements are not, however, generally valid in function space, i.e. in a space of infinitely many dimensions. Thus it is not possible to select a convergent sequence of functions from every infinite set of continuous functions $f(x)$ with bounded norm, and from the relation $\lim_{n\to\infty} Nf_n = 0$ for a sequence of continuous

[1] This follows from *Dini's theorem*: If a series of *positive* continuous functions $\sum_{\nu=1}^{\infty} f_\nu(t)$ converges to a continuous function $S(t)$ in a closed domain G, its convergence is uniform. A brief sketch of the proof follows: Let us write $S_n(t) = \sum_{\nu=1}^{n} f_\nu(t)$; $S(t) = S_n(t) + R_n(t)$. If our assertion were false there would exist a positive number α, a set of numbers $n_1, n_2, \cdots$ tending to infinity, and associated values $t_1, t_2, \cdots$ in G such that $R_{n_i}(t_i) \geq \alpha$, i.e. $S_{n_i}(t_i) \leq S(t_i) - \alpha$. We may assume that the values t_i converge to a limit t which is in G. Now let N be a fixed integer; then, for $n_i \geq N$, we have $S_{n_i}(t_i) \not\leq S_N(t_i)$, and thus $S_N(t_i) \leq S(t_i) - \alpha$. We now let i increase beyond all bounds and obtain, because of the assumed continuity of all functions, $S_N(t) \leq S(t) - \alpha$, which is impossible for sufficiently large N.

functions it does not follow that $\lim_{n\to\infty} f_n = 0$. Consider, for example, the functions

$$f_n(x) = 1 - n^2x^2 \qquad \text{for } x^2 \leq 1/n^2,$$
$$f_n(x) = 0 \qquad \text{for } x^2 \geq 1/n^2$$

in the interval $-1 \leq x \leq +1$. Every subset of this set converges to the function

$$f(x) = 0 \qquad \text{for } x \neq 0$$
$$f(x) = 1 \qquad \text{for } x = 0,$$

which is discontinuous at $x = 0$, even though $\lim_{n\to\infty} Nf_n = 0$.

Nevertheless, to carry through the analogy between vectors and functions, i.e. to preserve, in function space, both Weierstrass's principle of accumulation and the above convergence theorems, is a task that is essential for many purposes, especially for proofs of convergence and existence. There are two ways to approach this problem: First, the set of admissible functions may be widened by changing the concepts of integral and convergence; this method is used in the theory of Lebesgue but is not necessary for this book and will not be employed here.[1] The other approach, which we shall adopt, is to restrict the admissible functions to a set for which the principle of convergence is valid. The restriction we shall impose, beyond the requirement that each admissible function be continuous, is that the set of all admissible functions be *equicontinuous*.

Suppose we are dealing with functions of one independent variable x. The requirement of equicontinuity means that, for every positive ϵ, there exists a positive number $\delta = \delta(\epsilon)$, depending only on ϵ and not on the particular function $f(x)$ of the set, such that, if $|x_1 - x_2| < \delta(\epsilon)$, then $|f(x_1) - f(x_2)| < \epsilon$, provided x_1 and x_2 are within the postulated range of the independent variable. For example, all functions $f(x)$ for which $\int_a^b f'^2(x)\,dx \leq M$, where M is a fixed constant, form an equicontinuous set of functions in the interval $a \leq x \leq b$. Indeed, for any two values x_1 and x_2 in this interval,

$$f(x_2) - f(x_1) = \int_{x_1}^{x_2} f'(x)\,dx;$$

[1] See, however, §10, 11, of this chapter.

therefore, by the Schwarz inequality,

$$| f(x_2) - f(x_1) |^2 \leq | x_2 - x_1 | \int_{x_1}^{x_2} f'^2(x)\, dx \leq | x_2 - x_1 | M.$$

From this inequality we see that the condition for equicontinuity is satisfied for $\delta(\epsilon) = \epsilon^2/M$.

For equicontinuous sets of functions the *principle of accumulation* is valid: *From any set of functions which are uniformly bounded and equicontinuous in the fundamental domain G, it is possible to select a sequence $q_1(x)$, $q_2(x)$, $\cdots$ which converges uniformly to a continuous limit function in the domain G.*[1] This theorem (Theorem of Arzela) for sets of continuous functions is analogous to Weierstrass's theorem on bounded point sets; it therefore fulfills our requirement.

To prove the theorem we consider a denumerable set of points $x_1, x_2, \cdots$, which is everywhere dense in the interval, for example the set obtained by successive bisections of the interval and the resulting subintervals. The set of values of the functions at x_1 contains as a subset a convergent sequence; this follows from Weierstrass's theorem. Therefore it is possible to select from the given set of functions an infinite sequence of functions $a_1(x)$, $a_2(x)$, $\cdots$ whose values at the point x_1 form a convergent sequence. In the same way we can select from this sequence a subsequence $b_1(x)$, $b_2(x)$, $\cdots$ which converges at x_2, and so on. We now consider the "diagonal sequence" $a_1(x) = q_1(x)$, $b_2(x) = q_2(x)$, $\cdots$ and assert that it converges uniformly in the entire interval.

To show this we select an arbitrarily small positive number ϵ and a number M so large that for every point x of the interval there exists a point x_h with $h \leq M$ for which $| x - x_h | \leq \delta(\epsilon)$, $\delta(\epsilon)$ being the number defined above in the definition of equicontinuity. We now choose a number $N = N(\epsilon)$, depending only on ϵ, so great that, for $m > N$, $n > N$,

$$| q_m(x_h) - q_n(x_h) | < \epsilon \quad (h = 1, 2, \cdots, M).$$

[1] It is actually sufficient to assume uniform boundedness for a single point of G; because of the equicontinuity, the functions would then be uniformly bounded over the entire domain G.

Because of the equicontinuity of the functions we now have, for some particular value of $h \leq M$,

$$| q_m(x) - q_m(x_h) | < \epsilon,$$

$$| q_n(x) - q_n(x_h) | < \epsilon;$$

therefore, for $m > N$, $n > N$,

$$| q_m(x) - q_n(x) | < 3\epsilon.$$

This proves the uniform convergence of the sequence of functions $q_1(x)$, $q_2(x)$, $\cdots$ for all x in the interval $a \leq x \leq b$. The continuity of the limit function $q(x)$ follows immediately from the uniformity of the convergence. It may be remarked incidentally that these considerations also show that *every convergent subsequence converges uniformly.*

A set of equicontinuous functions posseses, moreover, the following properties: *If the sequence of functions* $f_1(x)$, $f_2(x)$, $\cdots$ *belongs to an equicontinuous set and if* $\lim_{n\to\infty} Nf_n = 0$, *then* $\lim_{n\to\infty} f_n = 0$. Furthermore, *if* f_n *is bounded and if* $\lim_{m\to\infty}^{n\to\infty} N(f_n - f_m) = 0$, *there exists a continuous function* $f(x)$ *such that* $\lim_{n\to\infty} f_n(x) = f(x)$. *In both cases the convergence to the limit is uniform.*

To prove the first assertion we assume that $\lim_{n\to\infty} f_n(x_0) \neq 0$ for some x_0. Then there exist arbitrarily large values of n such that $f_n^2(x_0) > 2\alpha^2$, where $2\alpha^2$ is some positive bound. Because of the equicontinuity of the functions $f_n(x)$ there must then exist a fixed interval of width δ, containing the point x_0, such that, within this interval, $f_n^2 > \alpha^2$ for the above values of n. Therefore $Nf_n > \delta\alpha^2$, in contradiction to the premise that $\lim_{n\to\infty} Nf_n = 0$. The proof of the second part of our assertion is similar; it may be carried out by the reader.

Another property of an equicontinuous set of functions with bounded norms is the following, which will be termed the smoothness[1] of the set: *Let* r *be a positive integer, and let* c_1, c_2, $\cdots$, c_r *be arbitrary numbers whose absolute values lie below a fixed bound, say* 1. *Then there exists a number* $\delta(\epsilon)$ *depending only on* r *and the positive number* ϵ *and tending to zero with* ϵ, *such that the relation*

$$| c_1f_1 + c_2f_2 + \cdots + c_rf_r | < \delta$$

[1] This concept, which relates to *sets* of functions, is not to be confused with the concept of a "smooth function" introduced at the beginning of this chapter.

follows from the relation $N(c_1f_1 + c_2f_2 + \cdots + c_rf_r) < \epsilon$, *if* $f_1, f_2, \cdots, f_r$ *are any* r *functions of the set.*

This follows from the theorems just proved if we note that the functions of our set remain equicontinuous when the set is extended by including all linear combinations $c_1f_1 + c_2f_2 + \cdots + c_rf_r$ with fixed r and bounded $|c_i|$.

The smoothness of the sequence of functions $f_1, f_2, \cdots$ may also be expressed in the following way: The sequence has the property that from each subsequence a uniformly convergent subsequence can be selected.

The following somewhat more general theorem may easily be derived from the principle of accumulation: *Let*

$$p_{11}(x), p_{12}(x), \cdots, p_{1r}(x),$$

$$p_{21}(x), p_{22}(x), \cdots, p_{2r}(x),$$

$$\cdots\cdots\cdots\cdots\cdots\cdots$$

be a sequence of sets $G_1, G_2, \cdots$ *of* r *functions each, such that all of the functions are equicontinuous and uniformly bounded in the interval* $a \leq x \leq b$. *Then a subsequence* $G_{n_1}, G_{n_2}, \cdots$ *may be selected such that the functions* $p_{n_i,k}(x)$ $(k = 1, 2, \cdots, r)$ *converge uniformly to* r *continuous functions* $p_1(x), p_2(x), \cdots, p_r(x)$ *as* i *increases.*

Selecting a suitable subsequence, the desired convergence may indeed be obtained for the functions of the first column; from this subsequence a new subsequence can be selected such that the second column also converges. This process is then repeated $r - 2$ more times.

§3. *Measure of Independence and Dimension Number*

1. Measure of Independence. A simple criterion for the linear dependence or independence of r functions $f_1, f_2, \cdots, f_r$ follows by a procedure analogous to that employed for vectors in n-dimensional space. We begin by considering the quadratic form in r real variables $t_1, t_2, \cdots, t_r$

$$K(t, t) = N(t_1f_1 + \cdots + t_rf_r) = \int (t_1f_1 + \cdots + t_rf_r)^2\, dx$$

(12)

$$= \sum_{i,k=1}^{r} (f_if_k)t_it_k.$$

Since $K(t, t)$ is positive definite, its lowest eigenvalue m (i.e. the minimum value of K when the t_i are varied subject to the restriction $\sum_{i=1}^{r} t_i^2 = 1$) is certainly not negative. We call this number the "*measure of independence*" of the functions $f_1, f_2, \cdots, f_r$. The functions $f_1, f_2, \cdots, f_r$ are evidently linearly dependent if and only if the measure of independence m is equal to zero. In the case of linear independence m indicates the "degree of independence".

The vanishing of the *Gram determinant*

$$\Gamma(f_1, \cdots, f_r) = \begin{vmatrix} (f_1 f_1) & \cdots & (f_1 f_r) \\ \cdots\cdots & \cdots & \cdots\cdots \\ (f_r f_1) & \cdots & (f_r f_r) \end{vmatrix} \tag{13}$$

of the system of functions $f_1, f_2, \cdots, f_r$ is equivalent to the vanishing of its measure of independence. This follows from the fact that the Gram determinant is the product of all the eigenvalues of K. Since, moreover, the eigenvalues are non-negative, we know that $m^r \leq \Gamma \leq mM^{r-1}$, where M is the greatest eigenvalue of K, and $\Gamma \geq 0$.[1] *Thus, the vanishing of the Gram determinant is a necessary and sufficient condition for the linear dependence of the functions* $f_1, f_2, \cdots, f_r$.

If we form a linear combination $f = \sum_{i=1}^{r} u_i f_i$ of r linearly independent functions $f_1, f_2, \cdots, f_r$ and if, moreover, f is normalized, we find that none of the coefficients u_i can exceed in absolute value the bound $1/\sqrt{m}$, which depends only on the measure of independence. For, if we write

$$v_i = \frac{u_i}{\sqrt{\sum_{i=1}^{r} u_i^2}},$$

we evidently have from the definition of m

$$m \leq \int \left(\sum_{i=1}^{r} v_i f_i \right)^2 dx = \frac{Nf}{\sum_{i=1}^{r} u_i^2} = \frac{1}{\sum_{i=1}^{r} u_i^2};$$

therefore $\sum_{i=1}^{r} u_i^2 \leq 1/m$. *If a system of r functions whose measure of independence lies above the positive bound* μ *is orthogonalized, i.e. if the functions of the set are replaced by suitable normalized linear com-*

[1] The latter inequality is a generalization of the Schwarz inequality; in fact it is the Schwarz inequality if $r = 2$.

binations of the functions, no coefficients with absolute values in excess of $1/\sqrt{\mu}$ *can occur.*

2. Asymptotic Dimension of a Sequence of Functions. The sequence of normalized functions (or, more generally, functions of bounded norm) $f_1, f_2, \cdots$ is said to be exactly r-dimensional if any $r + 1$ functions of the sequence are linearly dependent, and at the same time there exists at least one set of r functions of the sequence which are linearly independent. In this case every function of the sequence may be written as a linear combination $t_1g_1 + t_2g_2 + \cdots + t_rg_r$ with constant coefficients $t_1, t_2, \cdots, t_r$ of r basis functions $g_1, g_2, \cdots, g_r$; the entire sequence of functions is composed of members of the "linear space" (or "linear family") $t_1g_1 + t_2g_2 + \cdots + t_rg_r$.

If the sequence of functions $f_1, f_2, \cdots$ is not of finite dimension, two possibilities exist: 1) For every arbitrarily large positive integer s there exist sets of s functions $f_{n_1}, f_{n_2}, \cdots, f_{n_s}$ with arbitrarily large indices $n_1, n_2, \cdots, n_s$ such that the measure of independence of these functions lies above a fixed bound, independent of the indices n_i but possibly depending on s. In this case we ascribe the asymptotic dimension ∞ to the sequence of functions.[1] 2) For sufficiently large s the measure of independence of $f_{n_1}, \cdots, f_{n_s}$ converges to zero if all the indices $n_1, n_2, \cdots, n_s$ increase beyond all bounds. Here the smallest integer r for which the measure of independence tends to zero for $s > r$ is said to be the *asymptotic dimension* of the sequence. In particular, $r = 0$ if Nf_n tends to zero with increasing n. If the asymptotic dimension of a sequence is r and if sufficiently many functions at the beginning of the sequence are omitted, then any $r + 1$ of the remaining functions are "almost" linearly dependent.

This terminology, suggested by the analogy to vectors in n-dimensional space, would be meaningful if the functions sufficiently far out in a sequence of asymptotic dimension r could be approximated with prescribed accuracy by the functions of a linear space formed from r basis functions. In general, however, this is not true unless the concepts of function and of integral are extended according to the theory of Lebesgue. Since we wish to remain in the realm of our more elementary theory, we shall prove the above assertion by making certain restrictive assumptions as discussed in §2. In fact, we shall simply assume that the sequence of functions is *smooth.*

[1] The simplest example of a sequence with infinite dimension is a sequence of orthonormal functions for which the measure of independence of every finite subset has the value 1.

In this case, the following theorem holds: *If* $f_1, f_2, \cdots$ *is a smooth sequence of functions with asymptotic dimension* r, *then there exist* r *linearly independent functions (which may be taken to be orthonormal)* $g_1, g_2, \cdots, g_r$ *such that, for sufficiently large* n, *each function* f_n *differs from a function of the linear space* $t_1g_1 + t_2g_2 + \cdots + t_rg_r$ *by less than an arbitrarily small positive quantity* ϵ, *while no space with fewer than* r *basis functions has this property.*

This linear space may also be characterized in the following way: Let $G_1, G_2, \cdots, G_m, \cdots$ be sets of r functions $f_{m_1}, f_{m_2}, \cdots, f_{m_r}$ of the sequence, such that the measure of independence of each set lies above a fixed positive bound μ and the indices m_i $(i = 1, \cdots, r)$ increase beyond all bounds with increasing m. Then the linear spaces of functions S_m whose basis functions are the functions of G_m converge, with increasing m, uniformly to a limiting space T, defined by r linearly independent functions $g_1, g_2, \cdots, g_r$. By "converge" we mean here that, for sufficiently large m, every normalized function of S_m differs arbitrarily little from a function of T.

To formulate the proof conveniently, we introduce the concept of distance between a function and a space. We say that the distance between a function f and the linear space of functions S is less than the positive quantity d, if there is a function of S such that the absolute value of the difference between this function and f is everywhere less than d. Similarly we ascribe a distance less than d to two linear spaces of functions S and S^* if every normalized function of one space differs by an amount whose absolute value is less than d from some normalized function of the other space.

It now follows immediately that, for sufficiently large m and n, the distance between the function f_n and the space S_m is arbitrarily small. For, the measure of independence of $f_n, f_{m_1}, \cdots, f_{m_r}$ is certainly arbitrarily small for sufficiently large m and n. Since our sequence of functions is smooth, there exist $r + 1$ numbers $u_0, u_1, \cdots, u_r$, $\sum_{i=0}^{r} u_i^2 = 1$, for which $| u_0f_n + u_1f_{m_1} + \cdots + u_rf_{m_r} |$ is arbitrarily small. Note that the absolute value of u_0 cannot become arbitrarily small with increasing m and n, since otherwise the measure of independence of $f_{m_1}, f_{m_2}, \cdots, f_{m_r}$ would become arbitrarily small, in contradiction to our assumption that the measure of independence lies above the bound μ. We may therefore divide $u_0f_n + u_1f_{m_1} + \cdots + u_rf_{m_r}$ by u_0, putting $u_i/u_0 = -t_i$ and con-

clude that, for sufficiently large n and m, the function f_n differs arbitrarily little from a suitable function $t_1 f_{m_1} + t_2 f_{m_2} + \cdots + t_r f_{m_r}$ of the linear space S_m. Thus, for sufficiently large m and n, the distance between the spaces S_n and S_m is also arbitrarily small. Now, let ϵ be a positive number, later to be taken sufficiently small, and let $\epsilon_1, \epsilon_2, \cdots$ be a sequence of positive numbers with $\sum_{i=1}^{\infty} \epsilon_i = \epsilon$. Furthermore, let m_i be a positive integer such that, for $n > m_i$ and $m > m_i$, the distance between S_n and S_m is less than ϵ_i. Let us now take any r normalized functions $h_{11}, h_{12}, \cdots, h_{1r}$ of the space S_{m_1} and determine the normalized functions $h_{21}, h_{22}, \cdots, h_{2r}$ of the space S_{m_2} $(m_2 > m_1)$ such that $| h_{2i} - h_{1i} | < \epsilon_1$ (this is possible because the distance between S_{m_1} and S_{m_2} is less than ϵ_1). In the same way we determine a set of normalized functions $h_{31}, h_{32}, \cdots, h_{3r}$ of the space S_{m_3} $(m_3 > m_2)$ such that $| h_{3i} - h_{2i} | < \epsilon_2$, and so forth. Since $| h_{pi} - h_{qi} | < \epsilon_p + \cdots + \epsilon_{q-1}$ $(p < q)$ the sequence of functions h_{ni} $(i = 1, \cdots, r)$ with any fixed i converges uniformly to a limit function g_i, and $| g_i - h_{1i} | < \epsilon$. If ϵ is chosen sufficiently small, the functions $g_1, g_2, \cdots, g_r$ will have a nonvanishing measure of independence if the functions $h_{11}, h_{12}, \cdots, h_{1r}$ do; thus the g_i are linearly independent. These functions $g_1, g_2, \cdots, g_r$ evidently fulfill all our requirements.

§4. *Weierstrass's Approximation Theorem. Completeness of Powers and of Trigonometric Functions*

1. Weierstrass's Approximation Theorem. The most elementary example of a complete system of functions is given by the powers

$$1, x, x^2, x^3, \cdots .$$

They form a complete system of functions in every closed interval $a \leq x \leq b$; in fact, the following *approximation theorem of Weierstrass*[1] holds: *Any function which is continuous in the interval $a \leq x \leq b$ may be approximated uniformly by polynomials in this interval.*

This theorem asserts more than completeness; it asserts the possibility of uniform convergence rather than just convergence in the mean.

To prove this, we assume that the interval $a \leq x \leq b$ lies wholly

[1] K. T. W. Weierstrass, Über die analytische Darstellbarkeit sogenannter willkürlicher Funktionen reeller Argumente, Sitzungsber. K. Akad. Wiss. Berlin, 1885, pp. 633–639, 789–805; Werke, Vol. 3, pp. 1–37, Berlin, 1903.

in the interior of the interval $0 < x < 1$; thus, two numbers α and β may be found with $0 < \alpha < a < b < \beta < 1$. We may suppose that the function $f(x)$, which is by assumption continuous in the interval $a \leq x \leq b$, has been extended continuously to the entire interval $\alpha \leq x \leq \beta$.

Let us now consider the integral

$$J_n = \int_0^1 (1 - v^2)^n \, dv.$$

We see immediately that J_n converges to zero with increasing n. Now, if δ is a fixed number in the interval $0 < \delta < 1$, and if we set

$$J_n^* = \int_\delta^1 (1 - v^2)^n \, dv,$$

we assert that

$$\lim_{n \to \infty} \frac{J_n^*}{J_n} = 0,$$

which means that for sufficiently large n the integral from 0 to δ forms the dominant part of the whole integral from 0 to 1. In fact, for $n \geq 1$,

$$J_n > \int_0^1 (1 - v)^n \, dv = \frac{1}{n + 1},$$

$$J_n^* = \int_\delta^1 (1 - v^2)^n \, dv < (1 - \delta^2)^n (1 - \delta) < (1 - \delta^2)^n,$$

$$\frac{J_n^*}{J_n} < (n + 1)(1 - \delta^2)^n$$

and hence

$$\lim_{n \to \infty} \frac{J_n^*}{J_n} = 0.$$

We now assume $a \leq x \leq b$ and form the expressions

$$P_n(x) = \frac{\int_a^b f(u)[1 - (u - x)^2]^n \, du}{\int_{-1}^1 (1 - u^2)^n \, du} \qquad (n = 1, 2, \cdots),$$

which are polynomials in x of degree $2n$ whose coefficients are quotients of definite integrals. We shall show that they afford the desired approximation.

By making the substitution $u = v + x$ we find for the numerator

$$\int_\alpha^\beta f(u)[1 - (u - x)^2]^n \, du = \int_{\alpha-x}^{\beta-x} f(v + x)[1 - v^2]^n \, dv$$

$$= \int_{\alpha-x}^{-\delta} + \int_{-\delta}^{\delta} + \int_{\delta}^{\beta-x}$$

$$= I_1 + I_2 + I_3,$$

where the positive number δ in the interval $0 < \delta < 1$ will be suitably fixed later. The integral I_2 may be transformed to

$$I_2 = f(x) \int_{-\delta}^{\delta} (1 - v^2)^n \, dv + \int_{-\delta}^{\delta} [f(v + x) - f(x)](1 - v^2)^n \, dv$$

$$= 2f(x)(J_n - J_n^*) + \int_{-\delta}^{\delta} [f(v + x) - f(x)](1 - v^2)^n \, dv.$$

Because of the uniform continuity of $f(x)$ in the interval $\alpha \leq x \leq \beta$ it is possible, for arbitrarily small $\epsilon > 0$, to choose a $\delta = \delta(\epsilon)$ in the interval $0 < \delta < 1$, depending only on ϵ, such that, for $| v | \leq \delta$ and $a \leq x \leq b$, $| f(v + x) - f(x) | \leq \epsilon$. It then follows that

$$\left| \int_{-\delta}^{\delta} [f(v + x) - f(x)](1 - v^2)^n \, dv \right| \leq \epsilon \int_{-\delta}^{\delta} (1 - v^2)^n \, dv$$

$$< \epsilon \int_{-1}^{1} (1 - v^2)^n \, dv$$

$$= 2\epsilon J_n .$$

Furthermore, if M is the maximum of $| f(x) |$ for $\alpha \leq x \leq \beta$ we obtain

$$| I_1 | < M \int_{-1}^{-\delta} (1 - v^2)^n \, dv = MJ_n^*,$$

$$| I_3 | < M \int_{\delta}^{1} (1 - v^2)^n \, dv = MJ_n^*.$$

Therefore, since the denominator in $P_n(x)$ is equal to $2J_n$,

$$| P_n(x) - f(x) | < 2M \frac{J_n^*}{J_n} + \epsilon.$$

Since $\lim_{n\to\infty}(J_n^*/J_n) = 0$, the right side may be made less than 2ϵ by a suitable choice of n; thus $f(x)$ is indeed approximated by $P_n(x)$ uniformly in the interval $a \leq x \leq b$.

2. Extension to Functions of Several Variables. It may be shown in exactly the same manner that a function f of m variables $x_1, x_2, \cdots, x_m$, which is continuous for $a_i \leq x_i \leq b_i$ $(i = 1, 2, \cdots, m;$ $0 < \alpha_i < a_i < b_i < \beta_i < 1)$, may be approximated uniformly by the polynomials

$$P_n(x_1, \cdots, x_m) = \frac{\displaystyle\int_{\alpha_1}^{\beta_1}\cdots\int_{\alpha_m}^{\beta_m} f(u_1,\cdots,u_m)[1-(u_1-x_1)^2]^n\cdots[1-(u_m-x_m)^2]^n\,du_1\cdots du_m}{\left[\displaystyle\int_{-1}^{1}(1-u^2)^n\,du\right]^m}.$$

3. Simultaneous Approximation of Derivatives. A similar argument leads to the following general result: A function $f(x_1, x_2, \cdots, x_m)$ which, along with its derivatives up to the k-th order, is continuous in the closed region $a_i \leq x_i \leq b_i$, may be uniformly approximated by polynomials $P(x_1, x_2, \cdots, x_m)$ in such a way that the derivatives of f up to the k-th order are also approximated uniformly by the corresponding derivatives of the polynomials.

To prove this, we again assume that $0 < a_i < b_i < 1$ and that the function and its derivatives are extended continuously into a larger rectangular region $\alpha_i \leq x_i < \beta_i$ $(0 < \alpha_i < a_i < b_i < \beta_i < 1)$ in such a way that the function and its derivatives up to the $(k-1)$-st order vanish on the boundary of the larger region. Then the polynomials $P_n(x_1, x_2, \cdots, x_m)$ defined in the previous subsection yield the desired approximation. This may be shown very simply by differentiating with respect to x_i under the integral sign, replacing this differentiation by the equivalent differentiation with respect to u_i, and finally transforming the integral by integration by parts, making use of the assumed boundary conditions.

4. Completeness of the Trigonometric Functions. From § 4, 1 we can deduce the important fact that the orthonormal system of trigonometric functions

$$(14)\qquad \frac{1}{\sqrt{2\pi}}, \quad \frac{\cos x}{\sqrt{\pi}}, \quad \frac{\sin x}{\sqrt{\pi}}, \quad \frac{\cos 2x}{\sqrt{\pi}}, \quad \frac{\sin 2x}{\sqrt{\pi}}, \quad \cdots$$

is complete in the interval $-\pi \leq x \leq \pi$. The following more inclusive theorem can also be proved: *Every function $f(x)$ which is continuous in the interval $-\pi \leq x \leq \pi$ and for which $f(-\pi) = f(\pi)$ may be approximated uniformly by trigonometric polynomials*

$$\frac{\alpha_0}{2} + \sum_1^n (\alpha_\nu \cos \nu x + \beta_\nu \sin \nu x)$$

where α_ν and β_ν are constants.

To prove this, we write θ instead of x and consider a ξ, η-plane with the polar coordinates ρ and θ ($\xi = \rho \cos \theta$, $\eta = \rho \sin \theta$). The function

$$\varphi(\xi, \eta) = \rho f(\theta)$$

is then continuous in the entire ξ, η-plane and coincides with the given function $f(\theta)$ on the unit circle $\xi^2 + \eta^2 = 1$. According to Weierstrass's approximation theorem it may be approximated uniformly by polynomials in ξ and η in a square containing the unit circle. If we then set $\rho = 1$, we see that $f(\theta)$ may be approximated uniformly by polynomials in $\cos \theta$ and $\sin \theta$. But, by well-known formulas of trigonometry, every such polynomial may also be written in the above form

$$\frac{\alpha_0}{2} + \sum_1^n (\alpha_\nu \cos \nu x + \beta_\nu \sin \nu x).$$

A continuous function $f(x)$ which does not satisfy the periodicity condition $f(-\pi) = f(\pi)$ may be replaced by a continuous function $g(x)$ satisfying this condition, in such a way that the integral $\int_{-\pi}^{\pi} (f(x) - g(x))^2 \, dx$ is arbitrarily small. From this it follows that every continuous function can be approximated in the mean by trigonometric polynomials, and therefore that the trigonometric functions form a complete set.

§5. *Fourier Series*

1. Proof of the Fundamental Theorem. It follows from the considerations of §1 and from the orthogonality of the trigonometric functions that the best approximation in the mean of degree n is obtained by the so-called *Fourier polynomial*

$$s_n(x) = \tfrac{1}{2} a_0 + \sum_1^n (a_\nu \cos \nu x + b_\nu \sin \nu x),$$

with

$$a_\nu = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos \nu x\, dx \qquad (\nu = 1, 2, \cdots, n)$$

$$b_\nu = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin \nu x\, dx \qquad (\nu = 1, 2, \cdots, n) \tag{15}$$

$$a_0 = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\, dx.$$

This polynomial may also be written in the more concise form

$$s_n(x) = \sum_{\nu=-n}^{n} \alpha_\nu e^{i\nu x} \qquad \begin{cases} 2\alpha_\nu = a_\nu - ib_\nu, & \nu > 0, \\ 2\alpha_0 = a_0, & \\ 2\alpha_\nu = a_\nu + ib_\nu, & \nu < 0, \end{cases}$$

$$\alpha_\nu = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(t)\, e^{-i\nu t}\, dt \qquad (\nu = 0, \pm 1, \pm 2 \cdots) \tag{15'}$$

in virtue of the relation $\cos \nu x + i \sin \nu x = e^{i\nu x}$.

It is not a priori certain that these polynomials, which yield the best approximation in the mean, also yield a uniform approximation to the function—i.e. it is not certain that the infinite series $\lim_{n\to\infty} s_n(x)$ converges uniformly and represents the function $f(x)$. This question is the central problem of the theory of Fourier series.

For the sake of convenience we shall suppose that the function $f(x)$ is initially defined only in the interval $-\pi < x < \pi$, and then continued periodically beyond this interval by the functional equation $f(x + 2\pi) = f(x)$. Furthermore, at each jump discontinuity, we require $f(x)$ to be equal to the arithmetic mean of the "right-hand" and "left-hand" limits, $f(x + 0) = \lim_{h\to 0} f(x + h)$ and $f(x - 0) = \lim_{h\to 0} f(x - h)$ $(h > 0)$, respectively; i.e. we set $f(x) = \frac{1}{2}[f(x + 0) + f(x - 0)]$.

The following theorem then holds: *Every function which is piecewise smooth in the interval* $-\pi \leq x \leq \pi$ *and periodic with the period* 2π *may be expanded in a Fourier series*; that is, the Fourier polynomials

$$s_n(x) = \tfrac{1}{2} a_0 + \sum_{\nu=1}^{n} (a_\nu \cos \nu x + b_\nu \sin \nu x)$$

converge to $f(x)$ with increasing n. Moreover, we shall prove: *The*

convergence of the Fourier series is uniform in every closed interval in which the function is continuous.

We shall first give the proof for the case of continuous $f(x)$ in which discontinuities occur only in the derivative $f'(x)$. If the expansion coefficients of $f'(x)$ are denoted by α_ν and β_ν, we have

$$\alpha_\nu = \frac{1}{\pi}\int_{-\pi}^{\pi} f'(x)\cos \nu x\, dx = \frac{\nu}{\pi}\int_{-\pi}^{\pi} f(x)\sin \nu x\, dx = \nu b_\nu,$$

$$\beta_\nu = \frac{1}{\pi}\int_{-\pi}^{\pi} f'(x)\sin \nu x\, dx = -\frac{\nu}{\pi}\int_{-\pi}^{\pi} f(x)\cos \nu x\, dx = -\nu a_\nu,$$

$$\alpha_0 = 0.$$

Since $f'(x)$ is piecewise continuous, we have the completeness relation

$$\frac{1}{\pi}\int_{-\pi}^{\pi} f'^2(x)\, dx = \sum_{\nu=1}^{\infty}(\alpha_\nu^2 + \beta_\nu^2) = \sum_{\nu=1}^{\infty}\nu^2(a_\nu^2 + b_\nu^2).$$

Thus

$$\left|\sum_{\nu=n}^{m}(a_\nu \cos \nu x + b_\nu \sin \nu x)\right| = \left|\sum_{\nu=n}^{m}\frac{1}{\nu}(\nu a_\nu \cos \nu x + \nu b_\nu \sin \nu x)\right|$$

$$\leq \sqrt{\sum_{\nu=n}^{m}\nu^2(a_\nu^2 + b_\nu^2)}\cdot\sqrt{\sum_{\nu=n}^{m}\frac{1}{\nu^2}}$$

$$\leq \sqrt{\frac{1}{\pi}\int_{-\pi}^{\pi} f'^2\, dx}\cdot\sqrt{\sum_{\nu=n}^{m}\frac{1}{\nu^2}}.$$

This immediately establishes the absolute and uniform convergence of the infinite series

$$\lim_{n\to\infty} s_n(x) = \tfrac{1}{2}a_0 + \sum_{\nu=1}^{\infty}(a_\nu \cos \nu x + b_\nu \sin \nu x),$$

which represents the function $f(x)$, because of the completeness of the trigonometric functions.

To verify that the Fourier series expansion is also valid for functions which are discontinuous but piecewise smooth, we start by considering a particular function of this type, defined by

$$h(x) = \tfrac{1}{2}(\pi - x), \qquad 0 < x < 2\pi,$$

$$h(0) = 0,$$

$$h(x + 2\pi) = h(x).$$

This function jumps by π at the points $x = \pm 2k\pi$ $(k = 0, 1, \cdots)$. Its Fourier coefficients are

$$a_0 = 0, \qquad a_\nu = 0, \qquad b_\nu = \frac{1}{\nu} \qquad (\nu = 1, 2, \cdots).$$

If the Fourier expansion is valid for piecewise continuous functions we have

$$h(x) = \sum_{\nu=1}^{\infty} \frac{\sin \nu x}{\nu}.$$

To justify this equation we first form the function

$$g(x) = h(x)(1 - \cos x) = 2h(x) \sin^2 \frac{x}{2},$$

which is everywhere continuous and piecewise smooth. By the above reasoning, the Fourier series $\sum_{\nu=1}^{\infty} \beta_\nu \sin \nu x$ of this function converges uniformly and is equal to $g(x)$. The coefficients β_ν are related to the b_ν by the equations

$$\beta_\nu = b_\nu - \tfrac{1}{2}(b_{\nu-1} + b_{\nu+1}) \qquad (\nu = 2, 3, \cdots),$$
$$\beta_1 = b_1 - \tfrac{1}{2}b_2 .$$

If we set $\sum_{\nu=1}^{n} b_\nu \sin \nu x = s_n(x)$ and $\sum_{\nu=1}^{n} \beta_\nu \sin \nu x = \sigma_\nu(x)$ we have

$$(1 - \cos x)s_n(x) = \sigma_n(x) - \tfrac{1}{2}b_n \sin (n + 1)x + \tfrac{1}{2}b_{n+1} \sin nx.$$

With increasing n, b_n converges to zero and the sum $\sigma_n(x)$ converges uniformly to $g(x)$. It follows that $(1 - \cos x)s_n(x)$ also converges uniformly to $g(x)$ in the interval $-\pi \leq x \leq \pi$, and therefore that $s_n(x)$ itself converges uniformly to $h(x)$ in every closed subinterval not including the point $x = 0$.

At this excluded point $x = 0$ all partial sums s_n vanish, so that $\lim_{n\to\infty} s_n(0) = 0$. Thus the value of the series at the point of discontinuity is also equal to the value of $h(x)$, namely to the arithmetic mean of the right-hand and left-hand limits $+\pi/2$ and $-\pi/2$.

As the function $h(x)$ jumps by π at $x = 0$, the function $h(x - \xi)$ jumps by π at $x = \xi$ and is otherwise continuous in the basic interval. Now, if $f(x)$ is a piecewise continuous function which jumps by $s(\xi_i) = f(\xi_i + 0) - f(\xi_i - 0)$ at the points $x = \xi_i$ $(i = 1, \cdots, r)$ of the interval $0 \leq x < 2\pi$, then

$$F(x) = f(x) - \sum_{i=1}^{r} \frac{s(\xi_i)}{\pi} h(x - \xi_i)$$

is a function which is everywhere continuous and which, along with $f(x)$, has a piecewise continuous first derivative. Therefore, $F(x)$ may be expanded in a Fourier series which converges absolutely and uniformly. But the function

$$\sum_{i=1}^{r} \frac{s(\xi_i)}{\pi} h(x - \xi_i)$$

may also be expanded in a Fourier series, the convergence of which is uniform in every interval not containing any of the points of discontinuity. Thus the theorem formulated at the beginning of this section is proved completely.

2. Multiple Fourier Series. The trigonometric functions can be used to construct orthogonal systems over "cubes" in higher dimensional spaces. For simplicity we consider only the "plane" case (two dimensions). All our arguments will, however, be valid for any number of dimensions.

The functions

$$\cos \mu s \cos \nu t \qquad (\mu = 0, 1, \cdots ; \nu = 0, 1, \cdots),$$

$$\sin \mu s \cos \nu t \qquad (\mu = 1, 2, \cdots ; \nu = 0, 1, \cdots),$$

$$\cos \mu s \sin \nu t \qquad (\mu = 0, 1, \cdots ; \nu = 1, 2, \cdots),$$

$$\sin \mu s \sin \nu t \qquad (\mu = 1, 2, \cdots ; \nu = 1, 2, \cdots)$$

form an orthogonal system in the square $0 \leq s \leq 2\pi$, $0 \leq t \leq 2\pi$. The expansion formulas may be written most simply in terms of complex exponentials. If $F(s, t)$ can be expanded in a uniformly convergent double Fourier series, the series is

$$F(s, t) = \sum_{\mu=-\infty}^{\infty} \sum_{\nu=-\infty}^{\infty} a_{\mu\nu} e^{i(\mu s + \nu t)},$$

with

$$a_{\mu\nu} = \frac{1}{4\pi^2} \int_0^{2\pi} ds \int_0^{2\pi} dt \, F(s, t) \, e^{-i(\mu s + \nu t)}.$$

The completeness of the system of functions and, thus, the completeness relation

$$\sum_{\mu, \nu=-\infty}^{\infty} | a_{\mu\nu} |^2 = \int_0^{2\pi} \int_0^{2\pi} | F(s, t) |^2 \, ds \, dt$$

follow by our general theorem on the construction of complete systems in several variables from complete systems in one variable (see page 56).

Furthermore, the Fourier series for $F(s, t)$ converges absolutely and uniformly if $\partial^2 F(s, t)/\partial s\, \partial t$ exists and is piecewise continuous, (cf. subsection 1).

3. Order of Magnitude of Fourier Coefficients. If a periodic function $f(x)$ is continuous and has continuous derivatives up to the $(h - 1)$-st order, if its h-th derivative is piecewise continuous, and if the Fourier expansion of $f(x)$ is $f(x) = \sum_{\nu=-\infty}^{\infty} \alpha_\nu e^{i\nu x}$, then the coefficients α_ν, for $|\nu| \geq 1$, can be estimated by

$$2|\alpha_\nu| = \sqrt{a_\nu^2 + b_\nu^2} \leq \frac{c}{\nu^h},$$

where c is a constant. Thus, the smoother the function, the more rapidly the coefficients of the series tend to zero.

The above relation may be obtained immediately if expression (15′) for the coefficients is integrated by parts h times.

4. Change in Length of Basic Interval. If the function $f(x)$ is periodic with period $2l$ it may be expanded in the series

$$f(x) = \tfrac{1}{2}a_0 + \sum_{\nu=1}^{\infty} \left(a_\nu \cos \nu \frac{\pi}{l} x + b_\nu \sin \nu \frac{\pi}{l} x \right)$$

with

$$a_\nu = \frac{1}{l} \int_0^{2l} f(t) \cos \nu \frac{\pi}{l} t \, dt,$$

$$b_\nu = \frac{1}{l} \int_0^{2l} f(t) \sin \nu \frac{\pi}{l} t \, dt,$$

which may also be written in the form

$$f(x) = \sum_{\nu=-\infty}^{\infty} \alpha_\nu e^{i\nu(\pi/l)x},$$

$$a_\nu = \frac{1}{2l} \int_0^{2l} f(t) e^{-i\nu(\pi/l)t} \, dt.$$

5. Examples. For simple examples of the application of the theory of Fourier series the reader is referred to elementary texts.[1]

[1] For example R. Courant, Differential and Integral Calculus, Vol. 1, pp. 440–446, Blackie and Son, Ltd., London and Glasgow, 1934; 2nd ed. rev., Interscience Publishers, Inc., New York, 1937.

Here we shall employ the Fourier series expansion to derive the functional equation of the theta function and a general formula by Poisson.

The *functional equation of the theta function*

$$\theta(x) = \sum_{\mu=-\infty}^{\infty} e^{-\pi\mu^2 x} \qquad (x > 0)$$

is

$$\theta(x) = \frac{1}{\sqrt{x}}\,\theta\left(\frac{1}{x}\right).$$

To prove this relation, we set

$$\varphi(y) = \sum_{\mu=-\infty}^{\infty} e^{-\pi(\mu+y)^2 x};$$

$\varphi(y)$ is evidently a periodic function of y with the period 1, which possesses all derivatives with respect to y and which may therefore be expanded in the Fourier series

$$\varphi(y) = \sum_{\nu=-\infty}^{\infty} \alpha_\nu e^{2\pi i\nu y}$$

with

$$\alpha_\nu = \int_0^1 \varphi(t)\, e^{-2\pi i\nu t}\, dt = \int_0^1 \sum_{\mu=-\infty}^{\infty} e^{-\pi(\mu+t)^2 x - 2\pi i\nu t}\, dt.$$

Since the orders of summation and integration may be interchanged for all $x > 0$, we obtain for the coefficients α_ν :

$$\begin{aligned}
\alpha_\nu &= \sum_{\mu=-\infty}^{\infty} \int_0^1 e^{-\pi(\mu+t)^2 x - 2\pi i\nu(\mu+t)}\, dt \\
&= \sum_{\mu=-\infty}^{\infty} \int_{\mu}^{\mu+1} e^{-\pi t^2 x - 2\pi i\nu t}\, dt \\
&= \int_{-\infty}^{\infty} e^{-\pi t^2 x - 2\pi i\nu t}\, dt \\
&= e^{-(\pi\nu^2/x)} \int_{-\infty}^{\infty} e^{-\pi x(t+(i\nu/x))^2}\, dt = \frac{e^{-\pi\nu^2/x}}{\sqrt{x}}\ ;
\end{aligned}$$

for, $\int_{-\infty}^{\infty} e^{-z^2}\, dz$ has the same value $\sqrt{\pi}$ along a line $\mathfrak{I}m\ \mathrm{t} = \nu/x$ parallel

to the real axis as along the real axis itself. (If the Cauchy theorem is applied to the function e^{-z^2} and to the rectangle with vertices at $-T, +T, +T + i\nu/x, -T + i\nu/x$, and if T is then made to tend to infinity, the integrals over the vertical sides of the rectangle converge to zero; for, the integrand converges uniformly to zero, while the length of the path of integration is constant, equal to ν/x.) Thus we have

$$\varphi(y) = \frac{1}{\sqrt{x}} \sum_{\nu=-\infty}^{\infty} e^{-\pi\nu^2/x} e^{2\pi i\nu y}$$

and in particular, for $y = 0$,

$$\theta(x) = \sum_{\nu=-\infty}^{\infty} e^{-\pi\mu^2 x} = \varphi(0) = \frac{1}{\sqrt{x}} \sum_{\nu=-\infty}^{\infty} e^{-\pi\nu^2/x} = \frac{1}{\sqrt{x}}\theta\left(\frac{1}{x}\right).$$

Here the Fourier expansion has been applied to the transformation of infinite series in one special case. The method thus exemplified has recently proved very useful in the treatment of certain analytic functions arising in number theory.

This method is based on a very important transformation formula for infinite series, known as *Poisson's summation formula*. Let $\sum_{n=-\infty}^{\infty} \varphi(2\pi n)$ be an infinite series in which $\varphi(x)$ is a continuous and continuously differentiable function of x such that the series $\sum_{n=-\infty}^{\infty} \varphi(2\pi n + t)$ and $\sum_{n=-\infty}^{\infty} \varphi'(2\pi n + t)$ converge absolutely and uniformly for all t in the interval $0 \leq t < 2\pi$. The second series is then the derivative of the first, which may therefore be expanded in a convergent Fourier series in the interval $0 \leq t < 2\pi$:

$$\sum_{n=-\infty}^{\infty} \varphi(2\pi n + t) = \frac{1}{2\pi} \sum_{\nu=-\infty}^{\infty} e^{i\nu t} \int_0^{2\pi} e^{-i\nu\tau} \sum_{n=-\infty}^{\infty} \varphi(2\pi n + \tau)\, d\tau$$

$$= \frac{1}{2\pi} \sum_{\nu=-\infty}^{\infty} e^{i\nu t} \sum_{n=-\infty}^{\infty} \int_0^{2\pi} \varphi(2\pi n + \tau) e^{-i\nu\tau}\, d\tau.$$

The sum over n in the last expression may be transformed as follows:

$$\sum_{n=-\infty}^{\infty} \int_0^{2\pi} \varphi(2\pi n + \tau)\, e^{-i\nu\tau}\, d\tau = \sum_{n=-\infty}^{\infty} \int_{2\pi n}^{2\pi(n+1)} \varphi(\tau)\, e^{-i\nu\tau}\, d\tau$$

$$= \int_{-\infty}^{\infty} \varphi(\tau)\, e^{-i\nu\tau}\, d\tau.$$

Therefore

$$\sum_{n=-\infty}^{\infty} \varphi(2\pi n + t) = \frac{1}{2\pi} \sum_{\nu=-\infty}^{\infty} e^{i\nu t} \int_{-\infty}^{\infty} \varphi(\tau)\, e^{-i\nu\tau}\, d\tau.$$

If, finally, we set $t = 0$, we obtain

$$\sum_{n=-\infty}^{\infty} \varphi(2\pi n) = \frac{1}{2\pi} \sum_{\nu=-\infty}^{\infty} \int_{-\infty}^{\infty} \varphi(\tau)\, e^{-i\nu\tau}\, d\tau.$$

This is Poisson's formula. Clearly it is valid if all the integrals occurring in the formula exist, if $\sum_{n=-\infty}^{\infty} \varphi(2\pi n + t)$ converges uniformly in t for $0 \leq t < 2\pi$, and if this series represents a function which can be expanded in a Fourier series.

§6. *The Fourier Integral*

1. The Fundamental Theorem. Consider a function $f(x)$ which is represented by a Fourier series

$$f(x) = \sum_{\nu=-\infty}^{\infty} \alpha_\nu\, e^{i\nu(\pi/l)x}, \qquad \alpha_\nu = \frac{1}{2l} \int_{-l}^{l} f(t)\, e^{-i\nu(\pi/l)t}\, dt$$

in the interval $-l < x < l$. It seems desirable to let l go to ∞, since then it is no longer necessary to require that f be continued periodically; thus one may hope to obtain a representation for a non-periodic function defined for all real x. We shall continue to assume that $f(x)$ is piecewise smooth in every finite interval and that, at the discontinuities, the value of the function is the arithmetic mean of the right-hand and left-hand limits. We now add the further assumption that the integral $\int_{-\infty}^{\infty} |\, f(x)\, |\, dx$ exists.

If we set $\pi/l = \delta$, we obtain

$$f(x) = \frac{1}{2\pi} \sum_{\nu=-\infty}^{\infty} \delta \int_{-l}^{l} f(t)\, e^{-i\nu\delta(t-x)}\, dt.$$

Letting $l \to \infty$, and therefore $\delta \to 0$, we find that the formula

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} du \int_{-\infty}^{\infty} f(t)\, e^{-iu(t-x)}\, dt \tag{16}$$

appears plausible; it is correct if the passage to the limit can be

justified. For real functions $f(x)$ this may also be written in the form

$$f(x) = \frac{1}{\pi}\int_0^\infty du \int_{-\infty}^{\infty} f(t)\cos u(t-x)\,dt. \tag{17}$$

A strict proof of the validity of this "*Fourier integral formula*" is more easily obtained by a direct verification of equation (16) or (17) than by a justification of the passage to the limit.

We start with the integral formula

$$\lim_{v\to\infty}\frac{1}{\pi}\int_{-a}^{a} f(x+t)\frac{\sin vt}{t}\,dt = \tfrac{1}{2}[f(x+0)+f(x-0)] = f(x),$$

where a is an arbitrary positive number. This formula, which is valid for every piecewise smooth function, is due to Dirichlet.[1] From this formula it follows that

$$\begin{aligned}\pi f(x) &= \lim_{v\to\infty}\int_{-a}^{a} f(x+t)\,dt\int_0^v \cos ut\,du\\ &= \lim_{v\to\infty}\int_0^v du\int_{-a}^{a} f(x+t)\cos ut\,dt\\ &= \int_0^\infty du\int_{-a}^{a} f(x+t)\cos ut\,dt.\end{aligned}$$

We assert that the integration over t may be extended from $-\infty$ to ∞. For, if $A > a$, we have

$$\int_0^v\int_{-A}^{A} - \int_0^v\int_{-a}^{a} = \int_0^v\int_{-A}^{-a} + \int_0^v\int_a^A = \int_{-A}^{-a}\int_0^v + \int_a^A\int_0^v$$

whence, since $\int_{-\infty}^{\infty}|f(x)|\,dx = C$ exists by hypothesis,

$$\begin{aligned}\left|\int_0^v\int_{-A}^{A} - \int_0^v\int_{-a}^{a}\right| &\le \left|\int_{-A}^{-a} f(x+t)\frac{\sin vt}{t}\,dt\right| + \left|\int_a^A f(x+t)\frac{\sin vt}{t}\,dt\right|\\ &\le \frac{1}{a}\left(\int_{-A}^{-a}|f(x+t)|\,dt + \int_a^A |f(x+t)|\,dt\right)\\ &\le \frac{1}{a}\int_{-\infty}^{\infty}|f(x+t)|\,dt = \frac{C}{a}.\end{aligned}$$

[1] This formula is usually employed as the foundation of the theory of Fourier series as well. For a proof see textbooks on the calculus, e.g. Courant, Differential and Integral Calculus, Vol. 1, p. 450.

If we now keep v fixed and let A tend to infinity, it follows that

$$\left| \int_0^v \int_{-\infty}^{\infty} - \int_0^v \int_{-a}^{a} \right| \leq \frac{C}{a}$$

and, passing to the limit $v \to \infty$, we obtain

$$\left| \lim_{v \to \infty} \int_0^v \int_{-\infty}^{\infty} - \pi f(x) \right| \leq \frac{C}{a}.$$

The right-hand side may be made arbitrarily small by a suitable choice of a, and thus the desired formula (17) is proved.

Since $\int_{-\infty}^{\infty} f(t) \cos u(t - x)\, dt$ is an even function of u, the above equation may also be written in the form

$$\pi f(x) = \tfrac{1}{2} \int_{-\infty}^{\infty} du \int_{-\infty}^{\infty} f(t) \cos u(t - x)\, dt.$$

On the other hand, $\int_{-\infty}^{\infty} f(t) \sin u(t - x)\, dt$ is an odd function of u; therefore

$$0 = \tfrac{1}{2} i \int_{-\infty}^{\infty} du \int_{-\infty}^{\infty} f(t) \sin u(t - x)\, dt$$

provided the integral converges.[1] By subtracting the last equation from the previous one we obtain formula (16), valid wherever $f(x)$ is continuous.

2. Extension of the Result to Several Variables. By repeated application of formula (16) *analogous formulas are obtained for piecewise smooth functions of several variables*; these formulas are valid wherever the functions are continuous. Thus

$$4\pi^2 F(x_1, x_2) = \iiiint F(t_1, t_2)\, e^{-i[u_1(t_1 - x_1) + u_2(t_2 - x_2)]}\, dt_1 du_1 dt_2 du_2,$$

under the assumption that the integrals

$$\int_{-\infty}^{\infty} | F(t_1, x_2) |\, dt_1 \qquad \text{and} \qquad \int_{-\infty}^{\infty} | F(x_1, t_2) |\, dt_2$$

[1] This is the case for those values of x for which $f(x)$ is continuous. The integral diverges at points of discontinuity, as may easily be seen from the function

$$f(x) = 1 \text{ for } | x | \leq 1, \qquad f(x) = 0 \text{ for } | x | > 1.$$

exist. In general, for n variables, we have

$$(2\pi)^n F(x_1, \cdots, x_n) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} F(t_1, \cdots, t_n)\, e^{-i[u_1(t_1-x_1)+\ldots+u_n(t_n-x_n)]}\, dt_1 du_1 \cdots dt_n du_n$$

under analogous assumptions. The integrations are to be performed in the order in which the differentials are written in this formula.

3. Reciprocity Formulas. The Fourier integral theorem (16) takes on an especially elegant form if one sets

$$g(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(t)\, e^{-iut}\, dt;$$

for, it then states that the equations

$$g(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(t)\, e^{-iut}\, dt, \qquad f(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} g(u)\, e^{iut}\, du$$

follow from each other. If the left-hand sides are assumed to be known, these equations form a pair of so-called integral equations each of which is the solution of the other and which are mutually reciprocal. For even and odd functions we have the sets of real equations

$$g(u) = \sqrt{\frac{2}{\pi}} \int_0^{\infty} f(t) \cos ut\, dt, \qquad f(t) = \sqrt{\frac{2}{\pi}} \int_0^{\infty} g(u) \cos ut\, du$$

and

$$g(u) = \sqrt{\frac{2}{\pi}} \int_0^{\infty} f(t) \sin ut\, dt, \qquad f(t) = \sqrt{\frac{2}{\pi}} \int_0^{\infty} g(u) \sin ut\, du,$$

respectively.

The corresponding reciprocity formulas for functions of several variables are

$$f(x_1, \cdots, x_n) = \frac{1}{(\sqrt{2\pi})^n} \int \cdots \int g(\xi_1, \cdots, \xi_n)\, e^{i(\xi_1 x_1+\ldots+\xi_n x_n)}\, d\xi_1 \cdots d\xi_n,$$

$$g(\xi_1, \cdots, \xi_n) = \frac{1}{(\sqrt{2\pi})^n} \int \cdots \int f(x_1, \cdots, x_n)\, e^{-i(\xi_1 x_1+\ldots+\xi_n x_n)}\, dx_1 \cdots dx_n,$$

and are certainly valid if f and g have derivatives up to the $(n+1)$-st order in the entire space.

§7. *Examples of Fourier Integrals*

1. The Fourier integral formula (17)

$$f(x) = \frac{1}{\pi}\int_0^\infty du \int_{-\infty}^\infty f(t)\cos u(t-x)\,dt$$

$$= \frac{1}{\pi}\int_0^\infty \cos ux\,du \int_{-\infty}^\infty f(t)\cos ut\,dt$$

$$+ \frac{1}{\pi}\int_0^\infty \sin ux\,du \int_{-\infty}^\infty f(t)\sin ut\,dt$$

reduces to

$$f(x) = \frac{2}{\pi}\int_0^\infty \cos ux\,du \int_0^\infty f(t)\cos ut\,dt,$$

if $f(x)$ is an even function, and to

$$f(x) = \frac{2}{\pi}\int_0^\infty \sin ux\,du \int_0^\infty f(t)\sin ut\,dt$$

if $f(x)$ is odd.

2. We now consider *Dirichlet's discontinuous factor:* Let the even function $f(x)$ be defined by

$$f(x) = 1 \qquad \text{for } |x| < 1$$

$$f(x) = \tfrac{1}{2} \qquad \text{for } |x| = 1$$

$$f(x) = 0 \qquad \text{for } |x| > 1;$$

we may express it as the Fourier integral

$$f(x) = \frac{2}{\pi}\int_0^\infty \cos ux\,du \int_0^1 \cos ut\,dt$$

$$= \frac{2}{\pi}\int_0^\infty \frac{\sin u\cos ux}{u}\,du.$$

The term on the right, called "Dirichlet's discontinuous factor," is useful in many problems.

3. If, for $x > 0$, we take

$$f(x) = e^{-\beta x} \qquad (\beta > 0)$$

we have either

$$f(x) = \frac{2}{\pi}\int_0^\infty \cos ux\, du \int_0^\infty e^{-\beta t}\cos ut\, dt$$

$$= \frac{2}{\pi}\int_0^\infty \frac{\beta \cos ux}{\beta^2 + u^2}\, du$$

or

$$f(x) = \frac{2}{\pi}\int_0^\infty \sin ux\, du \int_0^\infty e^{-\beta t}\sin ut\, dt$$

$$= \frac{2}{\pi}\int_0^\infty \frac{u \sin ux}{\beta^2 + u^2}\, du;$$

the former is obtained if, for negative values of x, $f(x)$ is continued as an even function, the latter if it is continued as an odd function. In the second case we must put $f(0) = 0$. The integral

$$\int_0^\infty \frac{\cos ux}{\beta^2 + u^2}\, du = \frac{\pi}{2}\frac{e^{-\beta|x|}}{\beta} \qquad (\beta > 0)$$

is sometimes called the Laplace integral.

4. The function

$$f(x) = e^{-x^2/2}$$

provides a particularly instructive example. In this case, the mutually reciprocal integral equations

$$g(u) = \sqrt{\frac{2}{\pi}}\int_0^\infty f(t)\cos ut\, dt = \sqrt{\frac{2}{\pi}}\int_0^\infty e^{-t^2/2}\cos ut\, dt = e^{-u^2/2}$$

and

$$f(t) = \sqrt{\frac{2}{\pi}}\int_0^\infty g(u)\cos ut\, du = \sqrt{\frac{2}{\pi}}\int_0^\infty e^{-u^2/2}\cos ut\, du = e^{-t^2/2}$$

are actually identical.

§8. *Legendre Polynomials*

1. Construction of the Legendre Polynomials by Orthogonalization of the Powers 1, x, x^2, $\cdots$. If the powers 1, x, x^2, $\cdots$ are orthogonalized in a given basic interval, say $-1 \leq x \leq 1$, (see §1) we obtain a complete orthonormal system of functions which is in many ways

even simpler than the system of trigonometric functions. This procedure yields a sequence of orthonormal polynomials which are uniquely determined to within a factor ± 1. They become unique if we require that the coefficient of the highest power of x in each polynomial be positive

We assert that these polynomials are, except for constant factors (depending on n but not on x), identical with the polynomials

$$P_0(x) = 1, \qquad P_n(x) = \frac{1}{2^n n!} \frac{d^n(x^2 - 1)^n}{dx^n} \qquad (n = 1, 2, \cdots),$$

which are known as the *Legendre polynomials.*[1] Since it is easily seen that there can exist, apart from constant factors, only *one* set of orthogonal polynomials in which every degree is represented, it suffices to prove that P_n is a polynomial of the n-th degree and that the system of the polynomials $P_n(x)$ is indeed orthogonal. Now $P_n(x)$ is evidently a polynomial of the n-th degree; explicitly,

$$P_n(x) = \frac{1}{2^n n!} \sum_{\nu=0}^{n} (-1)^{n-\nu} \binom{n}{\nu} \frac{(2\nu)!}{(2\nu - n)!} x^{2\nu-n}$$

$$= \sum_{\nu=0}^{n} (-1)^{n-\nu} \frac{1 \cdot 3 \cdot 5 \cdots (2\nu - 1)}{(n - \nu)!(2\nu - n)!\, 2^{n-\nu}} x^{2\nu-n}.$$

The terms in negative powers are to be omitted; this comes about automatically if we put $(-r)! = \infty$ for all positive integers r. For even n, the first term is thus

$$(-1)^{n/2} \frac{1 \cdot 3 \cdot 5 \cdots (n - 1)}{2 \cdot 4 \cdot 6 \cdots n};$$

for odd n it is

$$(-1)^{(n-1)/2} \frac{1 \cdot 3 \cdot 5 \cdots n}{2 \cdot 4 \cdot 6 \cdots (n - 1)} x.$$

The first few Legendre polynomials are

$$P_0(x) = 1, \qquad P_1(x) = x, \qquad P_2(x) = \tfrac{3}{2}x^2 - \tfrac{1}{2},$$

$$P_3(x) = \tfrac{5}{2}x^3 - \tfrac{3}{2}x, \qquad P_4(x) = \tfrac{35}{8}x^4 - \tfrac{15}{4}x^2 + \tfrac{3}{8}.$$

[1] A. M. Legendre, Recherches sur l'attraction des sphéroides homogènes, Mém. math. phys. prés. a l'Acad. sc. par divers sav., Vol. 10, 1785, pp. 411–434; Recherches sur la figure des planètes, Mém. math. phys., Reg.de l'Acad. sc., 1784, 1787, pp. 370–389.

To prove that the $P_n(x)$ form an orthogonal system, we shall denote $(x^2 - 1)^n$ by $u_n(x)$. We then have, for every non-negative integer $m < n$,

$$\int_{-1}^{1} P_n(x)x^m \, dx = \frac{1}{2^n n!} \int_{-1}^{1} u_n^{(n)}(x)x^m \, dx = 0.$$

This may be demonstrated if we remove the factor x^m by repeated partial integration and note that all derivatives of $u_n(x)$ up to the $(n - 1)$-st vanish at the limits of the interval of integration. It follows that

$$\int_{-1}^{1} P_n(x)P_m(x) \, dx = 0 \qquad (m < n),$$

and thus that any two different Legendre polynomials are in fact orthogonal. In order to obtain the necessary normalization factors we now compute $\int_{-1}^{1} [u_n^{(n)}(x)]^2 \, dx$ by repeated partial integration:

$$\int_{-1}^{1} u_n^{(n)}u_n^{(n)} \, dx = -\int_{-1}^{1} u_n^{(n-1)}u_n^{(n+1)} \, dx = \int_{-1}^{1} u_n^{(n-2)}u_n^{(n+2)} \, dx = \cdots$$

$$= (-1)^n \int_{-1}^{1} u_n u_n^{(2n)} \, dx = (2n)! \int_{-1}^{1} (1 - x)^n(1 + x)^n \, dx.$$

Now

$$\int_{-1}^{1} (1 - x)^n(1 + x)^n \, dx = \frac{n}{n + 1} \int_{-1}^{1} (1 - x)^{n-1}(1 + x)^{n+1} \, dx = \cdots$$

$$= \frac{n(n - 1)\cdots 1}{(n + 1)(n + 2)\cdots(2n)} \int_{-1}^{1} (1 + x)^{2n} \, dx$$

$$= \frac{(n!)^2}{(2n)!(2n + 1)} 2^{2n+1}$$

and therefore

$$\int_{-1}^{1} P_n^2(x) \, dx = \frac{2}{2n + 1}.$$

The desired normalized polynomials are therefore

$$\varphi_n(x) = \sqrt{\frac{2n+1}{2}}\, P_n(x)$$

$$= \sqrt{\frac{2n+1}{2}}\, \frac{1}{2^n n!}\, \frac{d^n(x^2-1)^n}{dx^n} \qquad (n = 0, 1, 2, \cdots).$$

The Legendre polynomials $P_n(x)$ have the property that

$$P_n(1) = 1;$$

this is seen immediately if the n-th derivative of $(x-1)^n(x+1)^n$ is evaluated according to the product rule and x is set equal to 1 in the resulting expression.

2. The Generating Function. The Legendre polynomials are particularly important in potential theory, where they occur as the expansion coefficients of a "*generating function*." Consider two points in a plane, one at a unit distance from the origin and the other at a distance $u < 1$ from the origin, and suppose that the angle between their radius vectors is $\cos^{-1} x$. The distance between the two points is then $\sqrt{1-2ux+u^2}$. If we now expand the reciprocal of this distance in powers of u we find, by the binomial theorem,

$$\frac{1}{\sqrt{1-2ux+u^2}} = \sum_{n=0}^{\infty} Q_n(x)u^n \tag{18}$$

where $Q_n(x)$ is a polynomial of the n-th degree in x. The function on the left is said to "generate" the expansion coefficients $Q_n(x)$. We may show that $Q_n(x)$ is identical with the Legendre polynomial $P_n(x)$ either by evaluating $Q_n(x)$ explicitly by the binomial theorem or by demonstrating that the polynomials $Q_n(x)$ obey the same orthogonality relations as the $P_n(x)$. We choose the latter procedure.

From the definition (18) we immediately obtain

$$\frac{1}{\sqrt{1-2xu+u^2}}\,\frac{1}{\sqrt{1-2xv+v^2}} = \sum_{n,m=0}^{\infty} Q_n(x)Q_m(x)u^n v^m.$$

Integrating the left-hand side with respect to x from -1 to $+1$ we obtain, by an elementary calculation,

$$\frac{1}{\sqrt{uv}} \log \frac{1+\sqrt{uv}}{1-\sqrt{uv}} = \sum_{n=0}^{\infty} \frac{2}{2n+1}\, u^n v^n.$$

Integrating the right-hand side term by term and equating the coefficients of $u^n v^m$ to those in the expression for the left-hand side we find

$$\int_{-1}^{1} Q_n(x)Q_m(x)\,dx = \begin{cases} 0 & \text{for} \quad n \neq m \\ \dfrac{2}{2n+1} & \text{for} \quad n = m. \end{cases}$$

By setting $x = 1$ in equation (18) we find $Q_n(1) = 1$. This completes the proof of the identity of $Q_n(x)$ with $P_n(x)$.

3. Other Properties of the Legendre Polynomials. (a) *Recursion Formula.* Differentiating the generating function with respect to u we immediately obtain the following recursion formula connecting three successive Legendre polynomials:

$$(19) \quad (n+1)P_{n+1}(x) - (2n+1)xP_n(x) + nP_{n-1}(x) = 0.$$

(b) *Differential Equation.* The n-th Legendre polynomial

$$y(x) = \frac{1}{2^n n!}\frac{d^n}{dx^n}(x^2-1)^n$$

satisfies the linear homogeneous second order differential equation

$$(20) \qquad (x^2-1)y'' + 2xy' - n(n+1)y = 0$$

or

$$(20') \qquad [(x^2-1)y']' - n(n+1)y = 0.$$

This may be proved by $(n+1)$-fold differentiation of the equation $(x^2-1)u' = 2nxu$, with $u = (x^2-1)^n$ and $u^{(n)} = 2^n n! y$.[1]

(c) *Minimum Property.* If the Legendre polynomial $P_n(x)$ is multiplied by the reciprocal C of the coefficient of x^n, so that the coefficient of x^n in $CP_n(x)$ is equal to 1, we obtain polynomials characterized by the following minimum property: they have the smallest distance in the mean from zero of all polynomials of the n-th degree with the leading coefficient 1. To prove this we note that, in the integral $\int_{-1}^{1} (x^n + a_{n-1}x^{n-1} + \cdots + a_0)^2\,dx$, the integrand may be

[1] From this differential equation it follows that the zeros of $P_n(x)$ (all of which by Rolle's theorem are real and lie in the interval $-1 < x < 1$) are all simple, since the second and all higher derivatives of $P_n(x)$ would vanish at any multiple zero.

written in the form $(CP_n(x) + c_{n-1}P_{n-1}(x) + \cdots + c_0)^2$. The integral is therefore equal to

$$\frac{2C^2}{2n+1} + 2\sum_{\nu=1}^{n-1} \frac{c_\nu^2}{2\nu+1},$$

and this expression takes on its least value for

$$c_0 = c_1 = \cdots = c_{n-1} = 0.$$

§9. *Examples of Other Orthogonal Systems*

1. Generalization of the Problem Leading to Legendre Polynomials. We shall now generalize the problem which led to the definition of Legendre polynomials.

Let a non-negative "weight function" $p(x)$ be given in the interval $a \leq x \leq b$. The problem is to study the system of functions obtained by orthogonalizing $\sqrt{p(x)}$, $x\sqrt{p(x)}$, $x^2\sqrt{p(x)}$, $\cdots$ in the interval $a \leq x \leq b$.

These functions are, of course, linearly independent, as are the powers $1, x, x^2, \cdots$. In the orthogonal system the factors of $\sqrt{p(x)}$ are polynomials $Q_0(x)$, $Q_1(x)$, $\cdots$ of degree 0, 1, $\cdots$ which may be determined uniquely by normalizing conditions and which are termed "*orthogonal polynomials belonging to the weight function* $p(x)$."[1]

For example, for

$$a = -1, \qquad b = 1, \qquad p(x) = 1$$

we obtain the Legendre polynomials $P_n(x)$; for

$$a = -1, \qquad b = 1, \qquad p(x) = \frac{1}{\sqrt{1-x^2}},$$

[1] The polynomials $Q_0(x)$, $Q_1(x)$, $\cdots$, when multiplied by suitable factors C, possess a minimum property similar to that of the Legendre polynomials: the integral

$$\int p(x)(x^n + a_{n-1}x^{n-1} + \cdots + a_0)^2\,dx$$

takes on its least value when the polynomial in the integrand is $CQ_n(x)$. The polynomial in the integrand may again be written as a linear combination of the $Q_i(x)$, in the form $(CQ_n(x) + c_{n-1}Q_{n-1}(x) + \cdots + c_0)$. Since the functions $\sqrt{p(x)}Q_n(x)$ are orthogonal, and, in fact, orthornormal if the $Q_i(x)$ are appropriately defined, the integral is equal to $C^2 + \sum_{\nu=0}^{n-1} c_\nu^2$, which assumes its minimum at $c_0 = c_1 = \cdots = c_{n-1} = 0$.

the *Tchebycheff polynomials*

$$T_n(x) = \frac{1}{2^{n-1}} \cos (n \cos^{-1} x);$$

for

$$a = -1, \qquad b = 1, \qquad p(x) = \sqrt{1 - x^2},$$

the polynomials

$$Q_n(x) = \frac{\sin [(n + 1) \cos^{-1} x]}{\sqrt{1 - x^2}};$$

for

$$a = 0, \quad b = 1, \quad p(x) = x^{q-1}(1 - x)^{p-q} \qquad (q > 0,\ p - q > -1),$$

the *Jacobi* or *hypergeometric polynomials*; for

$$a = -\infty, \qquad b = \infty, \qquad p(x) = e^{-x^2},$$

the *Hermite polynomials*; for

$$a = 0, \qquad b = \infty, \qquad p(x) = e^{-x},$$

the *Laguerre polynomials*.

We shall now study the Tchebycheff, Jacobi, Hermite, and Laguerre polynomials in more detail.

2. **Tchebycheff Polynomials.**[1] The Tchebycheff polynomials

$$T_0(x) = 1, \qquad T_n(x) = \frac{1}{2^{n-1}} \cos (n \cos^{-1} x) \qquad (n \geq 1)$$

form an orthogonal system of polynomials[2] with the weight function $p(x) = 1/\sqrt{1 - x^2}$ in the interval $-1 \leq x \leq 1$, since

$$\int_{-1}^{1} T_n(x) T_m(x) \frac{dx}{\sqrt{1 - x^2}} = \frac{1}{2^{n+m-2}} \int_0^{\pi} \cos n\theta \cos m\theta \, d\theta = 0$$

for $n \neq m$.

[1] P. L. Tchebycheff, Sur les questions de minima, qui se rattachent à la représentation approximative des fonctions, Mém. Acad. sc. Pétersb., Ser. 6, Vol. 7, 1859, pp. 199–291; Oeuvres, Vol. 1, pp. 271–378, esp. pp. 295–301, St. Petersburg, 1899.

[2] These functions are indeed polynomials, since $\cos n\theta$ is a polynomial in $\cos \theta$:

$$\cos n\theta = \cos^n \theta - \binom{n}{2} \cos^{n-2} \theta \sin^2 \theta + \binom{n}{4} \cos^{n-4} \theta \sin^4 \theta - \cdots.$$

$T_n(x)$ is the polynomial of n-th degree with leading coefficient 1 whose maximum absolute value is smallest in the interval $-1 \leq x \leq 1$; i.e. it deviates least from zero. (The coefficient of x^n in T_n is equal to 1, as is easily seen.)

Proof: Let $\cos^{-1} x = \theta$, and consider the points $x_k = \cos k\pi/n$ $(k = 0, 1, \cdots, n)$, for which $T_n(x)$ attains its greatest deviation from zero. Clearly, for

$$\theta = 0, \quad \frac{\pi}{n}, \quad \frac{2\pi}{n}, \cdots, \pi,$$

we have

$$T_n(x) = \frac{1}{2^{n-1}}, \quad \frac{-1}{2^{n-1}}, \quad \frac{1}{2^{n-1}}, \cdots, \frac{(-1)^n}{2^{n-1}},$$

and, in general,

$$T_n(x_k) = \frac{(-1)^k}{2^{n-1}}.$$

Suppose the deviation from zero of a polynomial $R_n(x) = x^n + a_{n-1}x^{n-1} + \cdots$ in the interval $-1 \leq x \leq 1$ were less than that of $T_n(x)$. Then we would have

$$T_n(x_0) - R_n(x_0) > 0, \quad T_n(x_1) - R_n(x_1) < 0,$$
$$T_n(x_2) - R_n(x_2) > 0, \cdots;$$

i.e., the entire rational function $T_n(x) - R_n(x)$ would be alternately positive and negative at successive points x_k. This function would therefore have to have at least n roots; but this is impossible, since it is a polynomial of at most the $(n-1)$-st degree.

The polynomials $T_n(x)$ may be normalized if we divide by

$$\sqrt{\int_{-1}^{1} T_n^2 \frac{dx}{\sqrt{1-x^2}}} = \sqrt{\frac{\pi}{2^{2n-1}}}.$$

The Tchebycheff polynomials also occur as the expansion coefficients of the generating function

$$\psi(x, t) = \frac{1 - t^2}{1 - 2tx + t^2} = \sum_{n=0}^{\infty} T_n(x)(2t)^n. \tag{21}$$

Three successive Tchebycheff polynomials are connected by the recursion formula, valid for $n \geq 2$,

$$T_{n+1}(x) - xT_n(x) + \tfrac{1}{4}T_{n-1}(x) = 0. \tag{22}$$

For $n < 2$ the recursion formula takes a slightly different form:

$$T_2 - xT_1 + \tfrac{1}{4}T_0 = -\tfrac{1}{4},$$
$$T_1 - xT_0 = 0.$$

The Tchebycheff polynomials satisfy the linear homogeneous second order differential equation

$$(1 - x^2)y'' - xy' + n^2y = 0. \tag{23}$$

3. Jacobi Polynomials.[1] The Jacobi polynomials $G_n(p, q, x)$ are obtained for $a = 0, b = 1$ and the weight function

$$p(x) = x^{q-1}(1 - x)^{p-q} \qquad \text{with } q > 0, \quad p - q > -1.$$

They may also be obtained from the hypergeometric series

$$F(\alpha, \beta, \gamma, x) = 1 + \frac{\alpha}{1}\frac{\beta}{\gamma}x + \frac{\alpha(\alpha+1)}{1\cdot 2}\frac{\beta(\beta+1)}{\gamma(\gamma+1)}x^2 + \cdots \tag{24}$$

by replacing β by the negative integer $-n$, α by $p + n$, and γ by q. Therefore they satisfy the hypergeometric differential equation

$$x(1 - x)y'' + [\gamma - (\alpha + \beta + 1)x]y' - \alpha\beta y = 0 \tag{25}$$

or

$$x(1 - x)G_n''(x) + [q - (p + 1)x]G_n'(x) + (p + n)nG_n(x) = 0. \tag{25'}$$

They represent the sole entire rational solutions of this equation. The first few of these polynomials are

$$G_0(p, q, x) = 1,$$

$$G_1(p, q, x) = 1 - \binom{1}{1}\frac{p+1}{q}x,$$

$$G_2(p, q, x) = 1 - \binom{2}{1}\frac{p+2}{q}x + \binom{2}{2}\frac{(p+2)(p+3)}{q(q+1)}x^2,$$

$$G_3(p, q, x) = 1 - \binom{3}{1}\frac{p+3}{q}x + \binom{3}{2}\frac{(p+3)(p+4)}{q(q+1)}x^2 - \binom{3}{3}\frac{(p+3)(p+4)(p+5)}{q(q+1)(q+2)}x^3.$$

[1] C. G. J. Jacobi, Untersuchungen über die Differentialgleichung der hypergeometrischen Reihe, Journ. f. d. reine u. angew. Math., Vol. 56, 1859, pp. 149–165; Werke, Vol. 6, pp. 184–202, Berlin, 1891.

or, in general,

$$G_n(p, q, x) = \frac{x^{1-q}(1-x)^{q-p}}{q(q+1)\cdots(q+n-1)} \frac{d^n}{dx^n}[x^{q+n-1}(1-x)^{p+n-q}].$$

From this expression it can be seen that the Jacobi polynomials may also be defined in terms of a generating function by the relation

$$\frac{(1-x)^{1-q}(1+x)^{q-p}(t-1+\sqrt{1-2tx+t^2})^{q-1}(t+1-\sqrt{1-2tx+t^2})^{p-q}}{t^{p-1}\sqrt{1-2tx+t^2}} = \sum_{n=0}^{\infty} \binom{q+n-1}{n} G_n\left(p, q, \frac{1-x}{2}\right) t^n.$$

For $p = q = 1$ they reduce to the Legendre polynomials

$$(26) \qquad P_n(x) = G_n\left(1, 1, \frac{1-x}{2}\right) = F\left(n+1, -n, 1, \frac{1-x}{2}\right);$$

for $p = 0, q = \frac{1}{2}$ we obtain, essentially, the Tchebycheff polynomials

$$(27) \qquad T_n(x) = \frac{1}{2^{n-1}} G_n\left(0, \tfrac{1}{2}, \frac{1-x}{2}\right) = \frac{1}{2^{n-1}} F\left(n, -n, \tfrac{1}{2}, \frac{1-x}{2}\right).$$

4. Hermite Polynomials.[1] The Hermite polynomials $H_n(x)$ are the orthogonal polynomials in the interval $-\infty < x < \infty$ with the weight function $p(x) = e^{-x^2}$. They are most conveniently defined by means of a generating function $\psi(x, t)$:

$$(28) \qquad \psi(x, t) = e^{-t^2+2tx} = e^{x^2} e^{-(t-x)^2} = \sum_{n=0}^{\infty} \frac{H_n(x)}{n!} t^n.$$

From this equation it follows immediately that

$$(29) \qquad H_n(x) = \left(\frac{\partial^n \psi(x, t)}{\partial t^n}\right)_{t=0} = (-1)^n e^{x^2} \frac{d^n e^{-x^2}}{dx^n}.$$

The n-th Hermite polynomial $H_n(x)$ is thus $(-1)^n e^{x^2}$ times the n-th derivative of the function e^{-x^2}. From the relation $\partial\psi(x, t)/\partial x = 2t\psi(x, t)$ it follows that

$$(30) \qquad H_n'(x) = 2nH_{n-1}(x) \qquad (n \geq 1);$$

[1] C. Hermite, Sur un nouveau développement en série de fonctions, C. R. Acad. sc. Paris, Vol. 58, 1864, pp. 93–100, 266–273; Oeuvres, Vol. 2, pp. 293–312, Paris, 1908; Sur quelques développements en série de fonctions de plusieurs variables, C. R., Vol. 60, 1865, pp. 370–377, 432–440, 461–466, 512–518; Oeuvres, pp. 319–346.

from $\partial\psi(x, t)/\partial t + 2(t - x)\psi(x, t) = 0$ we obtain

$$H_{n+1}(x) - 2xH_n(x) + 2nH_{n-1}(x) = 0 \qquad (n \geq 1). \tag{31}$$

By combining (30) and (31) we obtain a linear homogeneous second order differential equation for $H_n(x)$:

$$H_n''(x) - 2xH_n'(x) + 2nH_n(x) = 0 \qquad (n \geq 0). \tag{32}$$

The first few Hermite polynomials are

$$H_0(x) = 1, \qquad H_1(x) = 2x,$$
$$H_2(x) = 4x^2 - 2, \qquad H_3(x) = 8x^3 - 12x,$$
$$H_4(x) = 16x^4 - 48x^2 + 12.$$

In general, the n-th Hermite polynomial is

$$H_n(x) = (2x)^n - \frac{n(n-1)}{1!}(2x)^{n-2} + \frac{n(n-1)(n-2)(n-3)}{2!}(2x)^{n-4} - \cdots.$$

The last term is

$$(-1)^{n/2}\frac{n!}{(n/2)!}$$

for even n and

$$(-1)^{(n-1)/2}\frac{n!}{((n-1)/2)!}2x$$

for odd n.

The orthogonality of the Hermite polynomials may be obtained[1] from

$$\int_{-\infty}^{\infty} H_m(x)H_n(x)\,e^{-x^2}\,dx = (-1)^n \int_{-\infty}^{\infty} H_m(x)\frac{d^n e^{-x^2}}{dx^n}\,dx$$

for $n > m$ by repeated partial integration, keeping in mind formula (30) and the fact that e^{-x^2} and all its derivatives vanish for infinite x:

$$\int_{-\infty}^{\infty} H_m(x)H_n(x)\,e^{-x^2}\,dx = (-1)^{n-1}\cdot 2m \int_{-\infty}^{\infty} H_{m-1}(x)\frac{d^{n-1}e^{-x^2}}{dx^{n-1}}\,dx = \cdots$$
$$= (-1)^{n-m}\,2^m\,m! \int_{-\infty}^{\infty} H_0(x)\frac{d^{n-m}e^{-x^2}}{dx^{n-m}}\,dx = 0.$$

[1] The orthogonality relations may also be obtained with the help of the generating function.

To obtain the normalization we set $n = m$ and have

$$\int_{-\infty}^{\infty} H_n^2(x)\, e^{-x^2}\, dx = 2^n n! \int_{-\infty}^{\infty} H_0(x) e^{-x^2}\, dx = 2^n n!\, \sqrt{\pi}.$$

Thus the functions of the orthonormal system are

$$\varphi_\nu(x) = \frac{H_\nu(x) e^{-x^2/2}}{\sqrt{2^\nu \nu!\, \sqrt{\pi}}} \qquad (\nu = 0, 1, 2, \cdots).$$

5. Laguerre Polynomials.[1] The Laguerre polynomial $L_n(x)$($a = 0$, $b = +\infty$, $p(x) = e^{-x}$) occurs as the factor of e^{-x} in the n-th derivative of the function $x^n e^{-x}$:

$$\begin{aligned} L_n(x) &= e^x \frac{d^n}{dx^n}(x^n e^{-x}) \\ &= \sum_{k=0}^{n} (-1)^k \binom{n}{k} n(n-1) \cdots (k+1) x^k \\ &= \sum_{k=0}^{n} (-1)^{n-k} \binom{n}{k} n(n-1) \cdots (n-k+1) x^{n-k} \\ &= \sum_{k=0}^{n} (-1)^{n-k} \frac{[n(n-1) \cdots (n-k+1)]^2}{k!} x^{n-k} \\ &= (-1)^n \left(x^n - \frac{n^2}{1!} x^{n-1} + \frac{n^2(n-1)^2}{2!} x^{n-2} - \cdots + (-1)^n n! \right); \end{aligned}$$

thus, for example,

$$\begin{aligned} L_0(x) &= 1, & L_1(x) &= -x + 1, \\ L_2(x) &= x^2 - 4x + 2, & L_3(x) &= -x^3 + 9x^2 - 18x + 6, \\ & & L_4(x) &= x^4 - 16x^3 + 72x^2 - 96x + 24. \end{aligned}$$

In virtue of the relations

$$\begin{aligned} \sum_{n=0}^{\infty} \frac{L_n(x)}{n!} t^n &= \sum_{n=0}^{\infty} \sum_{k=0}^{n} (-1)^k \binom{n}{k} \frac{1}{k!} x^k t^n \\ &= \sum_{k=0}^{\infty} \frac{(-1)^k x^k}{k!} \sum_{n=k}^{\infty} \binom{n}{k} t^n \\ &= \sum_{k=0}^{\infty} \frac{(-1)^k x^k}{k!} \frac{t^k}{(1-t)^{k+1}} = \frac{e^{-xt/(1-t)}}{(1-t)} \end{aligned}$$

[1] E. Laguerre, Sur l'intégrale $\int_x^\infty \frac{e^{-x}}{x}\, dx$, Bull. Soc. math. France, Vol. 7, 1879, pp. 72–81; Oeuvres, Vol. 1, pp. 428–437, Paris, 1898.

the Laguerre polynomials possess a simple generating function, namely

$$\psi(x, t) = \frac{e^{-xt/(1-t)}}{1-t}.$$

The relation

$$(1-t)^2 \frac{\partial \psi(x, t)}{\partial t} = (1 - t - x)\psi(x, t)$$

leads to the recursion formula

$$L_{n+1}(x) - (2n + 1 - x)L_n(x) + n^2 L_{n-1}(x) = 0 \quad (n \geq 1). \tag{33}$$

This, together with relation

$$L_n'(x) - nL_{n-1}'(x) = -nL_{n-1}(x) \qquad (n \geq 1), \tag{34}$$

which follows from the relation

$$(1-t)\frac{\partial \psi(x, t)}{\partial x} = -t\psi(x, t),$$

leads to the formula

$$xL_n'(x) = nL_n(x) - n^2 L_{n-1}(x) \qquad (n \geq 1) \tag{35}$$

and thus to the linear homogeneous second order differential equation

$$xy'' + (1 - x)y' + ny = 0 \tag{36}$$

satisfied by the Laguerre polynomial $L_n(x)$.

The orthogonality relation

$$\int_0^\infty e^{-x} L_n(x)L_m(x)\, dx = 0 \qquad (n > m)$$

follows from the equation

$$\begin{aligned}
\int_0^\infty e^{-x} x^k L_n(x)\, dx &= \int_0^\infty x^k \frac{d^n}{dx^n}(x^n e^{-x})\, dx \\
&= -k \int_0^\infty x^{k-1} \frac{d^{n-1}}{dx^{n-1}}(x^n e^{-x})\, dx \\
&= k(k-1) \int_0^\infty x^{k-2} \frac{d^{n-2}}{dx^{n-2}}(x^n e^{-x})\, dx = \cdots \\
&= (-1)^k k! \int_0^\infty \frac{d^{n-k}}{dx^{n-k}}(x^n e^{-x})\, dx = 0 \qquad \text{for } n > k,
\end{aligned}$$

and the normalization is given by

$$\int_0^\infty e^{-x} L_n^2(x)\, dx = \int_0^\infty (-1)^n x^n \frac{d^n}{dx^n}(x^n e^{-x})\, dx$$

$$= n! \int_0^\infty x^n e^{-x}\, dx = (n!)^2;$$

therefore the functions

$$\varphi_\nu(x) = \frac{e^{-x/2} L_\nu(x)}{\nu!} \qquad (\nu = 0, 1, 2, \cdots)$$

constitute the orthonormal system.[1]

6. Completeness of the Laguerre and Hermite Functions. The completeness of the Laguerre and Hermite functions remains to be investigated, since completeness has thus far been proved only for finite intervals. We call a system of functions in the interval $0 \leq x < \infty$ complete if every piecewise continuous function $f(x)$ for which the integral $\int_0^\infty f^2(x)\, dx$ exists can be approximated arbitrarily well in the mean by a linear combination of the functions.

To demonstrate the completeness of the Laguerre functions[2] we multiply both sides of the identity

$$\psi(x, t) = \frac{1}{1 - t} e^{-tx/(1-t)} = \sum_{n=0}^\infty \frac{t^n}{n!} L_n(x)$$

by $e^{-x/2}$, thus obtaining the corresponding identity

$$g(x, t) = \frac{1}{1 - t} e^{-\frac{1}{2}[(1+t)/(1-t)]x} = \sum_{n=0}^\infty t^n \varphi_n(x)$$

for the orthonormal Laguerre functions

$$\varphi_n(x) = e^{-x/2} \frac{L_n(x)}{n!}.$$

Now the infinite series $\sum_{n=0}^\infty t^n \varphi_n(x)$ converges in the mean to the generating function $g(x, t)$ for $|t| < 1$. This may easily be seen from the estimate

$$\int_0^\infty \left(g(x, t) - \sum_{n=0}^N t^n \varphi_n(x) \right)^2 dx = \frac{1}{1 - t^2} - \sum_{n=0}^N t^{2n},$$

[1] Once again the orthogonality relation could be obtained with the aid of the generating function.

[2] This proof was suggested in conversation by J. von Neumann.

obtained by means of the relations

$$\int_0^\infty g^2(x, t)\, dx = \frac{1}{1 - t^2},$$

and

$$\int_0^\infty g(x, t)\varphi_n(x)\, dx = t^n.$$

Since the quantity $\alpha = \frac{1}{2}(1 + t)/(1 - t)$ assumes all values from 0 to ∞ as t runs from -1 to $+1$, it follows that all functions of the form $e^{-\alpha x}$ can be approximated arbitrarily well in the mean by combinations of the Laguerre functions in the interval $0 \leq x < \infty$. Suppose that the function $f(x)$ is piecewise continuous and square-integrable in this interval. If we set $e^{-x} = \xi$, $f(x)$ goes over into a function $k(\xi)$ which is piecewise continuous in the interval $0 < \xi \leq 1$. The function $k(\xi)/\sqrt{\xi}$ is, moreover, square-integrable and can therefore be approximated in the mean by a function $G(\xi)$ which is piecewise continuous in the closed interval $0 \leq x \leq 1$. (For example, $G(\xi)$ can be taken identically equal to zero in a sufficiently small neighborhood of the origin and equal to $k(\xi)/\sqrt{\xi}$ elsewhere in the interval $0 \leq x \leq 1$.) The function $G(\xi)$ and, hence, $k(\xi)/\sqrt{\xi}$ can be approximated in the mean by polynomials

$$h_n(\xi) = a_0 + a_1\xi + a_2\xi^2 + \cdots + a_n\xi^n.$$

It follows that $f(x)$ can be approximated in the mean in the interval $0 \leq x < \infty$ by expressions of the form

$$\sqrt{\xi}\, h_n(\xi) = e^{-x/2}(a_0 + a_1e^{-x} + a_2e^{-2x} + \cdots + a_ne^{-nx})$$

and therefore by a combination of Laguerre functions. This fact is equivalent to the validity of the completeness relation

$$\sum_{\nu=0}^{\infty} c_\nu^2 = \int_0^\infty f^2(x)\, dx,$$

where c_ν denotes the expansion coefficient $c_\nu = \int_0^\infty f(x)\varphi_\nu(x)\, dx$.

The completeness of the Hermite polynomials can be proved on the basis of the completeness of the polynomials of Laguerre. Any square-integrable function $f(x)$ is the sum $f_1(x) + f_2(x)$ of an even function f_1 and an odd function f_2. For each of these functions the

completeness relation can be easily reduced to the completeness relation for Laguerre functions by substituting $x = u^2$. Details of the proof are omitted.

§10. *Supplement and Problems*

1. Hurwitz's Solution of the Isoperimetric Problem. The "isoperimetric problem" is the problem of finding the simple closed plane curve of given perimeter with maximum area. Its solution is known to be the circle. We confine ourselves to piecewise smooth curves. This problem is solved by Hurwitz[1] in the following way:

Let

$$x = x(s), \qquad y = y(s), \qquad 0 \leq s < L$$

be the parametric representation of a continuous piecewise smooth closed curve of perimeter L and area F. The parameter s is the arc length. We introduce instead the new parameter $t = 2\pi s/L$, which goes from 0 to 2π as s goes from 0 to L, and denote the Fourier coefficients of x and y by a_ν, b_ν and c_ν, d_ν, respectively; the Fourier coefficients of dx/dt, dy/dt are then νb_ν, $-\nu a_\nu$ and νd_ν, $-\nu c_\nu$. The relations

$$\left(\frac{dx}{ds}\right)^2 + \left(\frac{dy}{ds}\right)^2 = 1, \qquad \left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 = \left(\frac{L}{2\pi}\right)^2,$$

$$F = \int_0^{2\pi} x \frac{dy}{dt}\, dt$$

and the completeness relations (9) and (9′) then lead to

$$2\left(\frac{L}{2\pi}\right)^2 = \frac{1}{\pi}\int_0^{2\pi} \left\{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2\right\} dt = \sum_{\nu=1}^{\infty} \nu^2(a_\nu^2 + b_\nu^2 + c_\nu^2 + d_\nu^2),$$

$$\frac{F}{\pi} = \frac{1}{\pi}\int_0^{2\pi} x \frac{dy}{dt}\, dt = \sum_{\nu=1}^{\infty} \nu(a_\nu d_\nu - b_\nu c_\nu).$$

It follows that

$$L^2 - 4\pi F$$

$$= 2\pi^2 \sum_{\nu=1}^{\infty} [(\nu a_\nu - d_\nu)^2 + (\nu b_\nu + c_\nu)^2 + (\nu^2 - 1)(c_\nu^2 + d_\nu^2)] \geq 0.$$

[1] A. Hurwitz, Sur quelques applications géométriques des séries de Fourier, Ann. Éc. Norm., Ser. 3, Vol. 19, 1902, pp. 357–408, esp. pp. 392–397.

Evidently the equality can hold only if

$$b_1 + c_1 = 0, \qquad a_1 - d_1 = 0,$$

$$a_\nu = b_\nu = c_\nu = d_\nu = 0 \qquad \text{for } \nu = 2, 3, \cdots,$$

i.e. if

$$x = \tfrac{1}{2}a_0 + a_1 \cos t + b_1 \sin t,$$

$$y = \tfrac{1}{2}b_0 - b_1 \cos t + a_1 \sin t.$$

These are the equations of a circle. Thus all continuous piecewise smooth closed curves satisfy the "isoperimetric inequality"

$$L^2 - 4\pi F \geq 0, \tag{37}$$

where L is the perimeter and F the area. The equality holds if and only if the curve is a circle. This proves the isoperimetric character of the circle.

2. Reciprocity Formulas. Prove the equivalence of the two formulas[1]

$$f(t) = \int_0^1 g(u) \cot \pi(t - u)\, du$$

$$-g(u) = \int_0^1 f(t) \cot \pi(u - t)\, dt,$$

assuming that

$$\int_0^1 g(u)\, du = 0, \qquad \int_0^1 f(t)\, dt = 0,$$

and

$$g(u + 1) = g(u), \qquad f(t + 1) = f(t).$$

The integrals are to be taken as the "Cauchy principal values." The proof may be carried out by using either the theory of Fourier series or the Cauchy integral theorem.

3. The Fourier Integral and Convergence in the Mean. The theory of the Fourier integral may be developed in the same way as the theory of Fourier series in §5; again we use the concepts of convergence in the mean and completeness.

[1] Compare Hilbert, Integralgleichungen, p. 75.

Let the real or complex function $f(x)$ be piecewise continuous in every finite interval, and assume that the integrals

$$\int_{-\infty}^{\infty} | f(x) | \, dx, \qquad \int_{-\infty}^{\infty} | f(x) |^2 \, dx$$

exist. We attempt to find the best approximation in the mean of the function $f(x)$ by integrals of the form

$$\int_{-T}^{T} \varphi(t) \, e^{ixt} \, dt;$$

i.e. we attempt to make the integral

$$\int_{-\infty}^{\infty} \left| f(x) - \int_{-T}^{T} \varphi(t) \, e^{ixt} \, dt \right|^2 dx$$

as small as possible for a fixed value of T. It is not difficult to prove that

$$\int_{-\infty}^{\infty} \left| f(x) - \int_{-T}^{T} \varphi(t) \, e^{ixt} \, dt \right|^2 dx$$

$$= \int_{-\infty}^{\infty} | f(x) |^2 \, dx + 2\pi \int_{-T}^{T} | \varphi(x) - g(x) |^2 \, dx - 2\pi \int_{-T}^{T} | g(x) |^2 \, dx$$

with

$$g(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(\xi) e^{-ix\xi} \, d\xi.$$

Therefore our integral assumes its least value for

$$\varphi(t) = g(t).$$

Furthermore, the passage to the limit $T \to \infty$ yields the completeness relation

$$\int_{-\infty}^{\infty} | g(x) |^2 \, dx = \frac{1}{2\pi} \int_{-\infty}^{\infty} | f(x) |^2 \, dx.$$

This fact will not be proved here, nor shall we state the further considerations leading to the Fourier integral theorem itself.

4. Spectral Decomposition by Fourier Series and Integrals. Fourier series and Fourier integrals occur wherever one has occasion to represent a given phenomenon or function as a superposition of peri-

odic phenomena or functions. This representation is known as the *spectral decomposition* of the given function. If the Fourier series for the function $f(x)$ in the interval $-l \leq x < l$ is $\sum_{\nu=-\infty}^{\infty} \alpha_\nu e^{i\pi\nu x/l}$, the function $f(x)$ is said to be "decomposed" into periodic functions with the "discrete frequencies" $\nu\pi/l$ ($\nu = 0, 1, \cdots$) and the "amplitudes"

$$|\alpha_\nu| = \left| \frac{1}{2l} \int_{-l}^{l} f(x)\, e^{-i\pi\nu x/l}\, dx \right|.$$

If, on the other hand, the infinite interval $-\infty < x < \infty$ is considered, the function $f(x)$ is said to be decomposed into a "continuous spectrum," with the "spectral density"

$$g(u) = \int_{-\infty}^{\infty} f(x)\, e^{-iux}\, dx$$

at the frequency u.

An example of great interest in physics is given by the function[1]

$$f(x) = e^{i\omega x} \qquad \text{for } |x| < l$$

$$f(x) = 0 \qquad \text{for } |x| > l$$

which corresponds to a finite train of sinusoidal waves consisting of $n = l\omega/\pi$ waves. Its spectral density is given by

$$g(u) = \int_{-l}^{l} e^{i(\omega - u)x}\, dx = \frac{2 \sin (\omega - u)l}{\omega - u}.$$

The function $|g(u)|$ has a maximum at $u = \omega$, which becomes more pronounced as the number n of waves in the train becomes greater. Outside the arbitrarily small interval $\omega - \delta \leq u \leq \omega + \delta$, the spectral density becomes, in comparison with its maximum, arbitrarily small for large n.

5. Dense Systems of Functions. A system of functions will be called dense[2] if it has the property that every function $f(x)$ which

[1] This example is significant in optics; for, a sinusoidal wave train of finite length corresponds not to a sharp spectral line but to a spectrum of finite width which becomes sharper and more intense as the wave train becomes longer.

[2] Here we follow the definition of H. Müntz.

can be approximated arbitrarily well in the mean by a finite number of functions of the system can also be approximated in the mean by functions taken from any infinite subset of the original system of functions. The remarkable fact that nontrivial systems of this kind exist—a trivial system is, for example, one in which all the functions are equal—may be illustrated most easily on the basis of the following theorem:[1] *If* $\lambda_1, \lambda_2, \cdots, \lambda_n, \cdots$ *are positive numbers which tend to* ∞ *with increasing* n, *then the functions*

$$\frac{1}{x+\lambda_1}, \frac{1}{x+\lambda_2}, \cdots, \frac{1}{x+\lambda_n}, \cdots$$

form a complete system in every finite positive interval.

From this theorem it follows immediately that the system is dense, since every partial sequence of the λ_n also meets the specified requirements.

Because of the Weierstrass approximation theorem it suffices to show that every power x^m may be approximated uniformly by the functions $1/(x+\lambda_n)$. The rational function

$$x^m \frac{\lambda_p \lambda_{p+1} \cdots \lambda_q}{(x+\lambda_p)(x+\lambda_{p+1}) \cdots (x+\lambda_q)}$$

converges to x^m with increasing p and q $(q \geq p)$, and the convergence is uniform in every finite positive interval. If we always take $q - p \geq m$, this rational function may always be decomposed into partial fractions and written in the form

$$\frac{A_p}{x+\lambda_p} + \frac{A_{p+1}}{x+\lambda_{p+1}} + \cdots + \frac{A_q}{x+\lambda_q},$$

where $A_p, A_{p+1}, \cdots, A_q$ are constants, since all the numbers λ_n may be assumed to be different from each other. But this is a linear combination of the functions of the system under consideration.

Other examples of dense systems of functions have been given by H. Müntz.[2]

[1] G. Szegö, Über dichte Funktionenfamilien, Berichte der sächs. Akad. d. Wiss. zu Leipzig, Vol. 78, 1926, pp. 373–380.

[2] H. Müntz, Umkehrung bestimmter Integrale und absolute Approximation, Ch. II, Dichte Funktionensysteme, Math. Zeitschr., Vol. 21, 1924, pp. 104–110.

6. A Theorem of H. Müntz on the Completeness of Powers. Müntz[1] has proved the following interesting theorem: The infinite sequence of powers $1, x^{\lambda_1}, x^{\lambda_2}, \cdots$ with positive exponents which approach infinity is complete in the interval $0 \leq x \leq 1$ if and only if $\sum_{n=1}^{\infty} 1/\lambda_n$ diverges.

7. Fejér's Summation Theorem. From Weierstrass's approximation theorem we concluded that every continuous periodic function can be approximated uniformly by trigonometric polynomials. Such approximating polynomials may easily be constructed on the basis of the following theorem, which was discovered by Fejér.[2] *If $f(x)$ is a continuous periodic function and if $s_n(x)$ is the n-th partial sum of its Fourier series, then the sequence of arithmetic means*

$$S_n(x) = \frac{s_1(x) + s_2(x) + \cdots + s_n(x)}{n}$$

$$= \frac{2}{\pi n} \int_{-\pi}^{\pi} f(x + t) \left(\frac{\sin \frac{nt}{2}}{2 \sin \frac{t}{2}} \right)^2 dt$$

converges uniformly to $f(x)$.

An analogous theorem holds for the Fourier integral: Let $f(x)$ be continuous in every finite interval, and suppose that $\int_{-\infty}^{\infty} |f(x)|\, dx$ exists. Let us write

$$g(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(\xi)\, e^{-ix\xi}\, d\xi$$

and

$$s_T(x) = \int_{-T}^{T} g(t)\, e^{ixt}\, dt.$$

[1] H. Müntz, Über den Approximationssatz von Weierstrass, Festschr. H. A. Schwarz, p. 303, Julius Springer, Berlin, 1914; O. Szász, Über die Approximation stetiger Funktionen durch lineare Aggregate von Potenzen, Math. Ann., Vol. 77, 1916, pp. 482–496.

[2] L. Fejér, Untersuchungen über Fouriersche Reihen, Math. Ann., Vol. 58 1904, pp. 51–69.

Then the sequence of arithmetic means

$$S_T(x) = \frac{1}{T}\int_0^T s_T(x)\,dT = \frac{2}{\pi T}\int_{-\infty}^{\infty} f(x+t)\left(\frac{\sin\frac{Tt}{2}}{t}\right)^2 dt$$

converges uniformly to $f(x)$ in every finite interval. In particular, the convergence is uniform in the whole interval $-\infty < x < \infty$ if $f(x)$ is uniformly continuous there.

8. The Mellin Inversion Formulas.[1] THEOREM 1: Let $s = \sigma + ti$ be a complex variable. Let the function $f(s)$ be regular in the strip $\alpha < \sigma < \beta$ and let $\int_{-\infty}^{\infty} |f(\sigma + ti)|\,dt$ converge in this strip. Furthermore, let the function $f(s)$ tend uniformly to zero with increasing $|t|$ in every strip $\alpha + \delta \leq \sigma \leq \beta - \delta$ ($\delta > 0$, fixed arbitrarily). If for real positive x and fixed σ we define

$$g(x) = \frac{1}{2\pi i}\int_{\sigma-\infty i}^{\sigma+\infty i} x^{-s} f(s)\,ds, \tag{38}$$

then

$$f(s) = \int_0^{\infty} x^{s-1} g(x)\,dx \tag{39}$$

in the strip $\alpha < \sigma < \beta$.

Proof: Since, by assumption, $f(s)$ converges uniformly to zero for $\alpha + \delta \leq \sigma \leq \beta - \delta$ as $|t| \to \infty$, the path of integration in (38) may be displaced parallel to itself as long as it stays within the strip $\alpha < \sigma < \beta$; thus $g(x)$ does not depend on σ. If we choose two abscissas σ_1 and σ_2 with $\alpha < \sigma_1 < \sigma < \sigma_2 < \beta$, we have

$$\int_0^{\infty} x^{s-1} g(x)\,dx = \int_0^1 x^{s-1}\,dx\,\frac{1}{2\pi i}\int_{\sigma_1-\infty i}^{\sigma_1+\infty i} x^{-s_1} f(s_1)\,ds_1 + \int_1^{\infty} x^{s-1}\,dx\,\frac{1}{2\pi i}\int_{\sigma_2-\infty i}^{\sigma_2+\infty i} x^{-s_2} f(s_2)\,ds_2$$

$$= J_1 + J_2.$$

[1] H. Mellin, Über den Zusammenhang zwischen den linearen Differential- und Differenzengleichungen, Acta Math., Vol. 25, 1902, pp. 139–164, esp. 156–162; M. Fujiwara, Über Abelsche erzeugende Funktionen und Darstellbarkeitsbedingungen von Funktionen durch Dirichletsche Reihen, Tohoku math. J., Vol. 17, 1920, pp. 363–383, esp. 379–383; H. Hamburger, Über die Riemannsche Funktionalgleichung der ζ-Funktion (Erste Mitteilung), Math. Zeitschr., Vol. 10, 1921, pp. 240–254, esp. pp. 242–247.

The order of the integrations in these integrals may be interchanged, because we have the estimate

$$|J_1| \leq \frac{1}{2\pi}\int_{-\infty}^{\infty} |f(\sigma_1 + ti)|\,dt \int_0^1 x^{-1+(\sigma-\sigma_1)}\,dx < \infty,$$

$$|J_2| \leq \frac{1}{2\pi}\int_{-\infty}^{\infty} |f(\sigma_2 + ti)|\,dt \int_1^{\infty} x^{-1+(\sigma-\sigma_2)}\,dx < \infty$$

for the interchanged integrals.

We thus obtain

$$\int_0^{\infty} x^{s-1}g(x)\,dx = \frac{1}{2\pi i}\int_{\sigma_2-\infty i}^{\sigma_2+\infty i} \frac{f(s_2)}{s_2 - s}\,ds_2 - \frac{1}{2\pi i}\int_{\sigma_1-\infty i}^{\sigma_1+\infty i} \frac{f(s_1)}{s_1 - s}\,ds_1.$$

According to the Cauchy integral formula, the difference on the right side equals $f(s)$, since the integrals over horizontal segments connecting the two vertical lines $s = \sigma_1$ and $s = \sigma_2$ tend to zero for $|t| \to \infty$ (because $f(s) \to 0$).

THEOREM 2: Let $g(x)$ be piecewise smooth for $x > 0$, and let $\int_0^{\infty} x^{\sigma-1}g(x)\,dx$ be absolutely convergent for $\alpha < \sigma < \beta$. Then the inversion formula (38) follows from (39).

Proof: Let us put $x = e^u$. We then have

$$\frac{1}{2\pi i}\int_{\sigma-\infty i}^{\sigma+\infty i} x^{-s}f(s)\,ds = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-u(\sigma+ti)}\,dt \int_{-\infty}^{\infty} e^{v(\sigma+ti)}g(e^v)\,dv$$

$$= \frac{e^{-u\sigma}}{2\pi}\int_{-\infty}^{\infty} dt \int_{-\infty}^{\infty} e^{it(v-u)}e^{v\sigma}g(e^v)\,dv.$$

By the Fourier integral theorem (16) the latter expression is equal to $e^{-u\sigma}e^{u\sigma}g(e^u) = g(x)$; the theorem is therefore proved.

Examples of the Mellin integral transformation:

(a) Let

$$g(x) = \begin{cases} 1 \text{ for } 0 < x < 1 \\ \frac{1}{2} \text{ for } x = 1 \\ 0 \text{ for } x > 1; \end{cases}$$

since the integral $\int_0^{\infty} x^{\sigma-1}g(x)\,dx$ converges absolutely for $\sigma > 0$, we have

$$f(s) = \int_0^{\infty} x^{s-1}g(x)\,dx = \frac{1}{s} \qquad (\sigma > 0),$$

from which it follows that

$$g(x) = \frac{1}{2\pi i}\int_{\sigma-\infty i}^{\sigma+\infty i} \frac{x^{-s}}{s}\,ds \qquad (\sigma > 0).$$

This formula is important in the theory of Dirichlet series.

(b) From the definition of the Γ-function

$$\Gamma(s) = \int_0^\infty x^{s-1}e^{-x}\,dx \qquad (\sigma > 0)$$

one obtains

$$e^{-x} = \frac{1}{2\pi i}\int_{\sigma-\infty i}^{\sigma+\infty i} x^{-s}\Gamma(s)\,ds \qquad (\sigma > 0).$$

(c) The formula

$$\Gamma(s)\zeta(s) = \int_0^\infty \frac{x^{s-1}}{e^x - 1}\,dx \qquad (\sigma > 1)$$

where $\zeta(s)$ denotes the Riemann zeta function, leads to

$$\frac{1}{e^x - 1} = \frac{1}{2\pi i}\int_{\sigma-\infty i}^{\sigma+\infty i} x^{-s}\Gamma(s)\zeta(s)\,ds \qquad (\sigma > 1).$$

(d) The inversion of

$$\begin{aligned}\pi^{-s/2}\Gamma(s/2)\zeta(s) &= \int_0^\infty x^{s-1}\sum_{\nu=1}^{\infty} e^{-\pi\nu^2 x}\,dx \\ &= \int_0^\infty x^{s-1}\,\frac{\theta(x) - 1}{2}\,dx \qquad (\sigma > 1)\end{aligned}$$

is

$$\theta(x) = 1 + \frac{1}{\pi i}\int_{\sigma-\infty i}^{\sigma+\infty i} x^{-s}\pi^{-s/2}\Gamma(s/2)\zeta(s)\,ds \qquad (\sigma > 1).$$

The Mellin transformation is an important tool in the analytic theory of numbers and also occurs frequently in other problems of analysis.

9. The Gibbs Phenomenon. If we draw the graphs of a piecewise smooth function $f(x)$ and of its approximations by the partial sums of its Fourier series, then we find that the graphs of the partial sums approach the graph of $f(x)$ in every interval that does not contain a

discontinuity of $f(x)$. However, in the immediate vicinity of a jump discontinuity, where the convergence of the Fourier series is not uniform, the graphs of the partial sums contain oscillations which become progressively narrower and move closer to the point of discontinuity as the number of terms in the partial sums increases, but the total oscillation of the approximating curves does not approach the jump of $f(x)$. This is known as the *Gibbs phenomenon*.[1] To investigate it more closely we may, according to §5, confine ourselves to the particular Fourier series

$$\frac{\pi - x}{2} = \sum_{\nu=1}^{\infty} \frac{\sin \nu x}{\nu} \qquad (0 < x < 2\pi).$$

With the aid of the formula

$$s_n(x) = \sum_{\nu=1}^{n} \frac{\sin \nu x}{\nu} = -\frac{x}{2} + \int_0^x \frac{\sin(n + \frac{1}{2})t}{2 \sin \frac{1}{2} t}\, dt,$$

we may write the remainder of this series in the form

$$r_n(x) = \sum_{\nu=n+1}^{\infty} \frac{\sin \nu x}{\nu} = \frac{\pi}{2} - \int_0^x \frac{\sin (n + \frac{1}{2})t}{2 \sin \frac{1}{2}t}\, dt$$

or

$$r_n(x) = \frac{\pi}{2} - \int_0^{(n+\frac{1}{2})x} \frac{\sin t}{t}\, dt + \rho_n(x)$$

where

$$\rho_n(x) = \int_0^x \frac{2 \sin \frac{t}{2} - t}{2t \sin \frac{t}{2}} \sin (n + \tfrac{1}{2})t\, dt.$$

By differentiation we see easily that the approximation is worst at the points

$$x_k = \frac{2k\pi}{2n + 1} \qquad (k = 1, 2, \cdots, n),$$

[1] This fact was originally discovered empirically by Gibbs. J. W. Gibbs, Fourier's series, Nature, Vol. 59, 1898–99, pp. 200 and 606; Papers, Vol.2, pp. 258–260, Longmans, Green and Co., London, New York, and Bombay, 1906.

for which the remainder has maxima or minima. Its value at x_k is

$$r_n(x_k) = \frac{\pi}{2} - \int_0^{k\pi} \frac{\sin x}{x}\, dx + \rho_n\left(\frac{2k\pi}{2n+1}\right).$$

For increasing n and fixed k, $\rho_n(2k\pi/(2n+1))$ tends to zero. Thus the remainder $r_n(x_k)$, i.e. the deviation of the approximation from $\frac{1}{2}(\pi - x)$ at the point x_k (which approaches the point of discontinuity), tends to the limit

$$\lim_{n\to\infty} r_n(x_k) = \frac{\pi}{2} - \int_0^{k\pi} \frac{\sin x}{x}\, dx.$$

In particular $\lim r_n(x_1) \approx -0.2811$; i.e. the approximation curve overshoots the curve of $f(x)$ by about 9 percent of the amount of the jump discontinuity.[1]

It may be pointed out that the Gibbs phenomenon does not occur when Fejér's arithmetic means are used as the approximation functions.

10. A Theorem on Gram's Determinant. If G' is a subdomain of the basic domain G, if $\varphi_1, \varphi_2, \cdots, \varphi_n$ are piecewise smooth functions in G, and if Γ is their Gram determinant for the domain G and Γ' that for the domain G', then

$$\Gamma' \leq \Gamma.$$

The proof follows immediately from the maximum-minimum property of the eigenvalues. For, Γ is the product of the eigenvalues of the quadratic form

$$K(t, t) = \int_G (t_1\varphi_1 + \cdots + t_n\varphi_n)^2\, dG,$$

and Γ' is the corresponding product for

$$K'(t, t) = \int_{G'} (t_1\varphi_1 + \cdots + t_n\varphi_n)^2 dG.$$

[1] M. Bôcher, Introduction to the theory of Fourier's series, Ann. Math., Ser. 2, Vol. 7, 1906, pp. 81–152, esp. 123–132; C. Runge, Theorie und Praxis der Reihen, pp. 170–182, Göschen, Leipzig, 1904. For generalizations of the Gibbs phenomenon to other orthogonal systems, in particular systems in several variables, see H. Weyl, Die Gibbssche Erscheinung in der Theorie der Kugelfunktionen, Rend. Circ. mat. Palermo, Vol. 29, 1910, pp. 308–323; Über die Gibbssche Erscheinung und verwandte Konvergenzphänomene, ibid, Vol. 30, 1910, pp. 377–407.

Since, obviously,

$$K'(t, t) \leq K(t, t),$$

it follows immediately that every eigenvalue of $K'(t, t)$ is less than or equal to the corresponding eigenvalue of $K(t, t)$.

Another proof may be obtained from the following representation of the Gram determinant, in which we restrict ourselves, for the sake of brevity, to a single variable x in the basic domain $0 \leq x \leq 1$:

$$\Gamma = \left| \int_0^1 \varphi_i \varphi_k \, dx \right|$$

$$= \frac{1}{n!} \int_0^1 \int_0^1 \cdots \int_0^1 \begin{vmatrix} \varphi_1(x_1) & \varphi_1(x_2) & \cdots & \varphi_1(x_n) \\ \varphi_2(x_1) & \varphi_2(x_2) & \cdots & \varphi_2(x_n) \\ \cdots & \cdots & \cdots & \cdots \\ \varphi_n(x_1) & \varphi_n(x_2) & \cdots & \varphi_n(x_n) \end{vmatrix}^2 dx_1 \, dx_2 \cdots dx_n.$$

This representation corresponds exactly to formula (45) of Chapter I.[1]

11. Application of the Lebesgue Integral. Many topics discussed in this chapter become much more elegant if we use the Lebesgue integral instead of the more elementary Riemann integral. The set of admissible functions or "spaces" must be extended to include all functions which are integrable in the Lebesgue sense, and the methods of the Lebesgue theory must be applied.

The Lebesgue theory is based on the concept of the measure of a set of points $\mathfrak{M}$, all of which may be assumed to lie within a given finite interval. Let us suppose that the points of $\mathfrak{M}$ are all imbedded in a denumerable set of intervals; these intervals may, in general, overlap. Let m be the lower bound of the sum of the lengths of these intervals; moreover let m' be the corresponding lower bound for the points of the complementary set $\mathfrak{M}'$, i.e. for those points of the given interval that do not belong to $\mathfrak{M}$. If $m + m'$ equals the length of the given interval, the set $\mathfrak{M}$ is said to be *measurable* and m is said to be its *measure*. According to this definition the measure of every denumerable set is zero; such sets are called "null sets."

Consider a bounded function $f(x)$ defined in the interval

[1] Compare A. Kneser, Zur Theorie der Determinanten, Festschr. H. A. Schwarz, pp. 177–191, Julius Springer, Berlin, 1914; Gerhard W. H. Kowalewski, Einführung in die Determinantentheorie einschliesslich der unendlichen und der Fredholmschen Determinanten, Veit, Leipzig, 1909.

G ($a \leq x \leq b$), the values of which lie in an interval J. Let us subdivide J into subintervals $J_1, J_2, \cdots, J_n$. If for every such subinterval J_j the set of points of G for which $f(x)$ has values in J_j is measurable, then the function $f(x)$ is said to be *measurable* in G. In this case, if m_j denotes the measure of the set of points of G for which $f(x)$ has values in J_j, then the sum $\sum_{j=1}^{n} m_j f_j$ (f_j is any value of f lying in J_j) converges as the partition of J is refined, provided only that the lengths of the subintervals J_j tend to zero uniformly. The limit of the sum is called the *Lebesgue integral* or, simply, the integral of the function $f(x)$. It is a natural generalization of the ordinary Riemann integral and is denoted by the same symbol. The integral vanishes for a function which differs from zero only at the points of a null set. Thus a function may be changed arbitrarily in any null set, e.g. in the set of rational points, without affecting the value of its integral. Evidently, the domain of integrable functions has been greatly extended by this new definition. A function which is integrable in the sense of Lebesgue is said to be *summable.*

The concept of the Lebesgue integral may also be applied to functions which are not bounded in the given domain. We simply extend the integration first over those subdomains in which $-M < f(x) < N$, and then let M and N increase independently beyond all bounds. If the limit of the integral exists it is called the Lebesgue integral over the entire domain.

The following important theorems can be derived from these concepts:

(a) *Lebesgue Convergence Theorems.* If a sequence $f_1(x), f_2(x), \cdots$ of functions summable in the interval $a \leq x \leq b$ is given and if, for every x in the interval, the functions $f_n(x)$ converge with increasing n to a function $F(x)$, then the equation

$$\lim_{n\to\infty} \int f_n(x)\, dx = \int F(x)\, dx$$

is valid (even if the convergence is not uniform) provided that all the functions $f_n(x)$ have absolute values below a fixed bound independent of both n and x.

It is, in fact, sufficient that the inequality

$$|f_n(x)| < \varphi(x)$$

hold, where $\varphi(x)$ is a fixed summable function independent of n.

These theorems enable us to justify term-by-term integration of infinite series in many cases of nonuniform convergence.

(b) *Convergence in the Mean.* Suppose that the functions $f_1(x)$, $f_2(x)$, $\cdots$ and their squares are summable and that

$$\lim_{\substack{m\to\infty \\ n\to\infty}} \int (f_n - f_m)^2\, dx = 0.$$

The sequence of functions $f_n(x)$ is then said to *converge in the mean.* The following theorem holds: From any such sequence of functions it is possible to select a subsequence of f_{n_i} which converges pointwise to a summable function $f(x)$ except at points of a set of measure zero.

(c) *Theorem of Fischer and Riesz.*[1] This theorem may be formulated in two equivalent ways.

Fischer's formulation: Let the functions $f_1(x)$, $f_2(x)$, $\cdots$ and their squares be summable, and assume $\lim_{\substack{m\to\infty \\ n\to\infty}} \int (f_n - f_m)^2\, dx = 0$. Then there exists a square-summable function $f(x)$, such that

$$\lim_{n\to\infty} \int (f_n - f)^2\, dx = 0.$$

Riesz's formulation: If $\omega_1(x)$, $\omega_2(x)$, $\cdots$ is an arbitrarily given orthogonal system of functions and if a_1, a_2, $\cdots$ are arbitrary real numbers for which $\sum_{\nu=1}^{\infty} a_\nu^2$ converges, then there exists a summable function $f(x)$ with summable square for which $a_\nu = (f, \omega_\nu)$.

If the concepts of integral and function are generalized in the sense considered here then the above theorem establishes a biunique relationship between square-integrable functions and coordinates a_ν with a convergent sum of squares.

(d) *Completeness and Closedness of Systems of Functions.* We call a system of functions *closed* if there exists no normalized function orthogonal to all the functions of the system; we assume that these functions and their squares are summable. The following theorem holds: Every closed system of functions is complete and vice versa. In fact, if e.g. $f(x)$ is different from zero on a set of posi-

[1] F. Riesz, Sur les systèmes orthogonaux de fonctions, C. R. Acad. sc. Paris, Vol. 144, 1907, pp. 615–619; Über orthogonale Funktionensysteme, Nachr. Ges. Göttingen (math.-phys. Kl.) 1907, pp. 116–122; E. Fischer, Sur la convergence en moyenne, C. R. Acad. sc. Paris, Vol. 144, 1907, pp. 1022–1024.

tive measure and is at the same time orthogonal to all the functions of the orthogonal system, then $0 = \sum_{\nu=1}^{\infty} (f, \omega_\nu)^2 < \int f^2\, dx$; the system of functions is, therefore, not complete. On the other hand, if the system is not complete there exists a function $f(x)$ such that $\int f^2\, dx - \sum_{\nu=1}^{\infty} a_\nu^2 > 0$ for $a_\nu = (f, \omega_\nu)$; then, by Fischer's formulation of the Fischer-Riesz theorem, the functions $f_n = f - \sum_{\nu=1}^{n} a_\nu \omega_\nu$ converge in the mean to a function $\varphi(x)$ which is orthogonal to all the functions ω_ν. Therefore the system cannot be closed.

References

Textbooks

Borel, E., Leçons sur les fonctions de variables réelles et les développements en séries de polynomes. Gauthier-Villars, Paris, 1905.

Carslaw, H. S., Introduction to the theory of Fourier's series and integrals. 2nd edition. Macmillan, London, 1921.

Heine, E., Handbuch der Kugelfunktionen, 1 and 2. 2nd edition. G. Reimer, Berlin, 1878 and 1881.

Hilbert, D., Grundzüge einer allgemeinen Theorie der linearen Integralgleichungen. B. G. Teubner, Leipzig, 1912. (Cited as "Integralgleichungen.")

Hobson, E. W., The theory of functions of a real variable and the theory of Fourier's series. Cambridge University Press, Cambridge, 1907.

Lebesgue, H., Leçons sur l'intégration et la recherche des fonctions primitives. Gauthier-Villars, Paris, 1904. Leçons sur les séries trigonométriques. Gauthier-Villars, Paris, 1906.

Whittaker, E. T., and Watson, G. N., A course of modern analysis. 3rd edition. Cambridge University Press, Cambridge, 1920.

Monographs and Articles

Bôcher, M., Introduction to the theory of Fourier's series. Ann. Math., ser. 2, Vol. 7, 1906, pp. 81–152.

Courant, R., Über die Lösungen der Differentialgleichungen der Physik. Math. Ann., Vol. 85, 1922, pp. 280–325. Zur Theorie der linearen Integralgleichungen. Ibid., Vol. 89, 1923, pp. 161–178.

Hilbert, D., Über das Dirichletsche Prinzip. Festschr. Ges. Göttingen 1901, Berlin 1901; reprinted Math. Ann., Vol. 59, 1904, pp. 161–186.

Montel, P., Sur les suites infinies de fonctions. Ann. Éc. Norm., ser. 3, Vol. 24, 1907, pp. 233–334.

Szegö, G., Beitrag zur Theorie der Polynome von Laguerre und Jacobi. Math. Zeitschr., Vol. 1, 1918, pp. 341–356. Über Orthogonalsysteme von Polynomen. Ibid., Vol. 4, 1919, pp. 139–157. Über die Entwicklung einer willkürlichen Funktion nach den Polynomen eines Orthogonalsystems. Ibid., Vol. 12, 1922, pp. 61–94.

CHAPTER III

Linear Integral Equations

§1. *Introduction*

1. Notation and Basic Concepts. Let $K(s, t)$ be a function of the two variables s and t defined and continuous in the region $a \leq s \leq b$, $a \leq t \leq b$, and let λ be a parameter. Furthermore, let $f(s)$ and $\varphi(s)$ be two functions of the variable s continuous in the interval $a \leq s \leq b$, which are connected by the functional equation

$$f(s) = \varphi(s) - \lambda \int K(s, t)\, \varphi(t)\, dt. \tag{1}$$

(All integrals are to be extended over the "fundamental domain" as defined above, unless the contrary is explicitly stated.) The functional equation (1) is called a *linear integral equation of the second kind* with the kernel $K(s, t)$. By (1) every continuous function $\varphi(s)$ is transformed into another continuous function $f(s)$; the transformation is linear, since to $c_1\varphi_1 + c_2\varphi_2$ there corresponds the analogous combination $c_1 f_1 + c_2 f_2$. We shall be primarily concerned with the problem of solving the integral equation, i.e. with the problem of determining $\varphi(s)$ when $f(s)$ is given or of inverting the *linear integral transformation* (1). Unless the contrary is explicitly stated, we shall assume that all quantities considered are real.

If the function $f(s)$ vanishes identically we are dealing with a *homogeneous integral equation*. If a homogeneous equation possesses a solution other than the trivial solution $\varphi = 0$, the solution may be multiplied by an arbitrary constant factor and may therefore be assumed normalized. If $\varphi_1, \varphi_2, \cdots, \varphi_h$ are solutions of the homogeneous equation then all linear combinations $c_1\varphi_1 + c_2\varphi_2 + \cdots + c_h\varphi_h$ are also solutions. Therefore, if several linearly independent solutions are given we may assume that they are normalized and orthogonal; for, if they were not they could be orthogonalized by the procedure of Ch. II, §1 without ceasing to be solutions. We

shall assume that linearly independent solutions of the same homogeneous integral equation are orthonormal. A value λ (possibly complex) for which the homogeneous equation possesses nonvanishing solutions is called an *eigenvalue*[1] of the kernel; the corresponding solutions $\varphi_1, \varphi_2, \cdots, \varphi_h$ (assumed normalized and mutually orthogonal) are called the *eigenfunctions* of the kernel for the eigenvalue λ. Their number is finite for each eigenvalue. For, applying the Bessel inequality (Ch. II, §1) to the kernel $K(s, t)$ and the orthonormal functions $\varphi_1, \varphi_2, \cdots \varphi_h$, we find

$$\lambda^2 \int [K(s, t)]^2 \, dt \geq \lambda^2 \sum_{i=1}^{h} \left[\int K(s, t) \, \varphi_i(t) \, dt \right]^2 = \sum_{i=1}^{h} \varphi_i(s)^2;$$

by integrating with respect to s we obtain

$$\lambda^2 \iint [K(s, t)]^2 \, ds \, dt \geq h.$$

This establishes an upper bound for h. Thus *every eigenvalue possesses a finite multiplicity* (i.e. number of linearly independent eigenfunctions).

We shall see in §6 that the integral equation represents a generalization of the problem of linear algebra which was treated in Ch. I, §2. Its significance lies in the fact that it enables us to consider a variety of different problems in mathematical analysis from a unified point of view.

2. Functions in Integral Representation. In connection with equation (1), it is natural to inquire into the properties of functions that can be represented by an integral of the form

$$g(s) = \int K(s, t) h(t) \, dt. \tag{2}$$

In (2), $g(s)$ is called an *integral transform* of $h(s)$.

If $h(t)$ is piecewise continuous, $g(s)$ is certainly continuous. Moreover, if $\int [h(t)]^2 \, dt \leq M$, where M is a fixed bound, then the aggregate of functions defined by (2) is actually equicontinuous; i.e. for every positive ϵ there exists a positive number $\delta(\epsilon)$ independent of the particular function h such that $| g(s + \eta) - g(s) | < \epsilon$ whenever

[1] Such eigenvalues actually correspond to the reciprocals of the quantities called eigenvalues in Chapter I.

$|\eta| < \delta$ (compare Ch. II, §2). From the Schwarz inequality we have

$$[g(s+\eta) - g(s)]^2 \leq M \int [K(s+\eta, t) - K(s, t)]^2 \, dt$$

and since the kernel is uniformly continuous our assertion is proved; for, the inequality

$$|K(s+\eta, t) - K(s, t)| < \sigma$$

holds independent of t for arbitrarily small σ if η is sufficiently small.

Furthermore, if a sequence of kernels K_n is given for which

$$\lim K_n(s, t) = K(s, t)$$

uniformly, then for a given $h(t)$ the relation

$$g(s) = \lim_{n\to\infty} \int K_n(s, t)h(t) \, dt$$

holds, and the convergence is uniform in s, since the passage to the limit may be performed under the integral sign. It therefore follows that all functions of the form

$$g_n(s) = \int K_n(s, t)h(t) \, dt, \qquad g(s) = \int K(s, t)h(t) \, dt$$

for the functions $h(t)$ considered form an equicontinuous set as long as h is restricted by $\int h^2 \, dt \leq M$. These functions are also uniformly bounded, i.e. their absolute values lie below a common bound. This follows from the Schwarz inequality

$$[g_n(s)]^2 \leq M \int [K_n(s, t)]^2 \, dt; \qquad [g(s)]^2 \leq M \int [K(s, t)]^2 \, dt.$$

3. Degenerate Kernels. A kernel, which can be written as a finite sum of products of functions of s and functions of t

$$A(s, t) = \sum_{i=1}^{p} \alpha_i(s)\beta_i(t), \tag{3}$$

is called a *degenerate kernel.* Here we may assume that the functions $\alpha_i(s)$ are linearly independent of each other, and that the same is true of the functions $\beta_i(t)$. For, otherwise, one of these functions could be written as a linear combination of the others and $A(s, t)$ as a sum of fewer than p terms of the above form. The possibility of uniform approximation of a continuous function $K(s, t)$ by polynomials (Ch. II, §4) implies that the kernel $K(s, t)$ may be uniformly approximated by a degenerate kernel, since every polynomial in s and t evidently represents a degenerate kernel.

A degenerate kernel $A(s, t)$ may be transformed into another form, which is often more convenient. We suppose that the $2p$ functions of s: $\alpha_1(s)$, $\alpha_2(s)$, $\cdots$, $\alpha_p(s)$; $\beta_1(s)$, $\beta_2(s)$, $\cdots$, $\beta_p(s)$ are represented as linear combinations of a system of q orthonormal functions $\omega_1(s)$, $\omega_2(s)$, $\cdots$, $\omega_q(s)$. This can always be done by orthogonalization of the given functions. Then $A(s, t)$ appears in the form of a double sum

$$A(s, t) = \sum_{i,j=1}^{q} c_{ij}\omega_i(s)\omega_j(t). \tag{4}$$

The products $\omega_i(s)\omega_j(t)$ form a system of q^2 functions of s and t in the square $a \leq s \leq b$, $a \leq t \leq b$ which are mutually orthogonal and therefore linearly independent. If $A(s, t)$ is *symmetric,* i.e. if $A(s, t) = A(t, s)$ identically, then

$$\sum_{i,j=1}^{q} (c_{ij} - c_{ji})\omega_i(s)\omega_j(t) = 0;$$

because of the linear independence of the products $\omega_i(s)\omega_j(t)$ this implies $c_{ij} = c_{ji}$.

A symmetric kernel $K(s, t)$ can always be approximated uniformly by symmetric degenerate kernels $A(s, t)$. To see this we need only replace $A(s, t)$ by $\frac{1}{2}[A(s, t) + A(t, s)]$, which approximates the symmetric kernel $K(s, t)$ if $A(s, t)$ does.

§2. *Fredholm's Theorems for Degenerate Kernels*

The basic theorems of the general theory of integral equations, which were first proved by Fredholm,[1] correspond to the basic

[1] I. Fredholm, Sur une classe d'équations fonctionnelles, Acta Math., Vol. 27, 1903, pp. 365–390.

theorems of the theory of linear equations. They may be formulated in the following way:

Either the integral equation (1),

$$f(s) = \varphi(s) - \lambda \int K(s, t)\varphi(t)\, dt,$$

with fixed λ *possesses one and only one continuous solution* $\varphi(s)$ *for each arbitrary continuous function* $f(s)$, *in particular the solution* $\varphi = 0$ *for* $f = 0$; *or the associated homogeneous equation*

$$\psi(s) = \lambda \int K(s, t)\psi(t)\, dt \tag{5}$$

possesses a finite positive number r *of linearly independent solutions* $\psi_1, \psi_2, \cdots, \psi_r$. *In the first case the "transposed" integral equation*

$$g(s) = \varphi(s) - \lambda \int K(t, s)\varphi(t)\, dt \tag{6}$$

associated with (1) *also possesses a unique solution for every* g. *In the second case the transposed homogeneous equation*

$$\chi(s) = \lambda \int K(t, s)\chi(t)\, dt \tag{7}$$

also has r *linearly independent solutions* $\chi_1, \chi_2, \cdots, \chi_r$; *the inhomogeneous integral equation* (1) *has a solution if and only if the given function* $f(s)$ *satisfies the* r *conditions*

$$(f, \chi_i) = \int f(s)\chi_i(s)\, ds = 0 \quad (i = 1, 2, \cdots, r). \tag{8}$$

In this case the solution of (1) *is determined only up to an additive linear combination* $c_1\psi_1 + c_2\psi_2 + \cdots + c_r\psi_r$; *it may be determined uniquely by the additional requirements:*

$$(\varphi, \psi_i) = \int \varphi(s)\psi_i(s)\, ds = 0 \quad (i = 1, 2, \cdots, r).$$

We shall first prove these theorems for the case of the degenerate kernel $K(s, t) = A(s, t)$ represented by equation (3). Here the theory of our integral equation reduces almost immediately to that of a system of p linear equations in p unknowns. Let us write the integral equation in the form

$$f(s) = \varphi(s) - \lambda \sum_{i=1}^{p} \alpha_i(s) \int \beta_i(t)\varphi(t)\, dt. \tag{9}$$

Now setting $x_i = (\beta_i, \varphi)$, multiplying (9) by $\beta_j(s)$, and integrating with respect to s, we obtain the system of equations

$$f_j = x_j - \lambda \sum_{i=1}^{p} c_{ji} x_i \qquad (j = 1, 2, \cdots, p) \tag{10}$$

for the quantities x_i, where $f_j = (\beta_j, f)$ and $c_{ji} = (\beta_j, \alpha_i)$. If this system of equations possesses one and only one solution $x_1, x_2, \cdots, x_p$, then the function

$$\varphi(s) = f(s) + \lambda \sum_{i=1}^{p} x_i \alpha_i(s)$$

is certainly a solution of the integral equation, as may be verified by substituting in the integral equation and making use of equations (10). Then the transposed system of equations

$$g_j = y_j - \lambda \sum_{i=1}^{p} c_{ij} y_i \tag{11}$$

also possesses a unique solution $y_1, y_2, \cdots, y_p$, and

$$\varphi(s) = g(s) + \lambda \sum_{i=1}^{p} y_i \beta_i(s)$$

is a solution of the transposed integral equation (6).

If, on the other hand, the homogeneous system of equations

$$0 = x_j - \lambda \sum_{i=1}^{p} c_{ji} x_i \qquad (j = 1, 2, \cdots, p) \tag{12}$$

possesses a nontrivial solution $x_1, x_2, \cdots, x_p$, then

$$\psi(s) = \lambda \sum_{i=1}^{p} x_i \alpha_i(s)$$

is a nontrivial solution of the homogeneous integral equation (5). Two linearly independent solutions $x_1, x_2, \cdots, x_p$ and $x_1', x_2', \cdots, x_p'$ of (12) evidently yield two linearly independent solutions $\psi(s) = \lambda \sum_{i=1}^{p} x_i \alpha_i(s)$ and $\psi'(s) = \lambda \sum_{i=1}^{p} x_i' \alpha_i(s)$ of (5) and conversely.

The existence of r linearly independent solutions $\psi_1, \psi_2, \cdots, \psi_r$ of (5) and therefore of r independent systems of solutions of (12) is equivalent to the existence of the same number of linearly independent solutions $y_{k1}, y_{k2}, \cdots, y_{kp}$ $(k = 1, 2, \cdots, r)$ of the transposed system

$$g_j = y_j - \lambda \sum_{i=1}^{p} c_{ij} y_i \qquad (j = 1, 2, \cdots, p) \tag{13}$$

for $g_j = 0$, and therefore of r linearly independent solutions

$$\chi_1(s), \chi_2(s), \cdots, \chi_r(s)$$

of the transposed homogeneous integral equation (7); here

$$\chi_k(s) = \lambda \sum_{j=1}^{p} y_{kj}\beta_j(s). \tag{14}$$

According to the theory of linear equations, the equations (10), and therefore (13) and (6), always have unique solutions for $r = 0$; however, for $r > 0$, the inhomogeneous equation (10) and thus the integral equation (5) can be solved if and only if the conditions

$$\sum_{j=1}^{p} f_j y_{kj} = 0 \qquad (k = 1, 2, \cdots, r) \tag{15}$$

are satisfied. Because of the definitions of y_{kj} and f_j these conditions are equivalent to

$$(f, \chi_k) = 0 \qquad (k = 1, 2, \cdots, r). \tag{16}$$

This completes the proof of the Fredholm theorems for the case of degenerate kernels.

§3. *Fredholm's Theorems for Arbitrary Kernels*

In order to utilize the results of the previous section for the treatment of the integral equation with an arbitrary kernel $K(s, t)$, we employ the convergence theorem of Ch. II, §2.

We suppose that $K(s, t)$ is approximated uniformly by a sequence $A_1(s, t), A_2(s, t), \cdots, A_n(s, t), \cdots$ of degenerate kernels and consider, in addition to the integral equation (1), the approximating integral equations

$$f(s) = \varphi(s) - \lambda \int A_n(s, t)\varphi(t)\, dt. \tag{17}$$

If λ is fixed there are two possibilities:

Case I. For infinitely many indices n the equation (17) possesses a solution $\rho_n(s)$ for every $f(s)$, and for all these solutions $(\rho_n, \rho_n) = c_n^2 \leq M$, where M is a bound independent of n. In this case we may discard all unsuitable values of n and renumber the kernels A_n; without loss of generality we may assume that the condition is satisfied for all n.

Case II. The above statement is not true. Then, for some particular $f(s)$, either

(a) a solution $\rho_n(s)$ exists for infinitely many n (we may assume, as above, that it exists for all n), but $(\rho_n, \rho_n) = c_n^2 \to \infty$ as $n \to \infty$, or

(b) a solution exists only for a finite number of values of n (we may assume that it exists for no n), and therefore—by the Fredholm theorems for degenerate kernels—the homogeneous integral equation

$$0 = \varphi(s) - \lambda \int A_n(s, t)\varphi(t)\, dt \tag{18}$$

possesses a normalized solution $\sigma_n(s)$ for every n.

In Case I the functions $\rho_n(s) - f(s)$ are integral transforms of ρ_n by kernels $A_n(s, t)$; according to §1 they form an equicontinuous and uniformly bounded set. Therefore, by the convergence theorem of Ch. II, §2, there exists a subsequence of the functions $\rho_n(s)$ which converges uniformly to a limit function $\varphi(s)$. By passing to the limit in the integral equation (17), which is permissible, we see that this limit function satisfies the integral equation (1). In Case I this equation has, therefore, a solution for every $f(s)$.

In Case II (a) we divide the integral equation (17) for $\varphi = \rho_n$ by c_n and set $\rho_n/c_n = \sigma_n$, obtaining the equation

$$\frac{f(s)}{c_n} = \sigma_n(s) - \lambda \int A_n(s, t)\sigma_n(t)\, dt;$$

in Case II (b) we note that equation (18) holds for $\varphi = \sigma_n$. In both cases σ_n is normalized; hence the respective sequences of integral transforms $\sigma_n(s) - f(s)/c_n$ and $\sigma_n(s)$ are equicontinuous and uniformly bounded. Therefore subsequences from these sets converge uniformly to limit functions $\psi(s)$, which necessarily satisfy the homogeneous integral equation (5),

$$\psi(s) = \lambda \int K(s, t)\psi(t)\, dt,$$

and are normalized. Thus in Case II the homogeneous equation possesses nontrivial solutions, which we call null solutions or eigenfunctions.

To derive from these considerations the Fredholm theorems as formulated in §2 we recall the fact, proved in §1, that for every value

of λ only a finite number r of linearly independent eigenfunctions can exist. If $r = 0$, Case II, which always leads to a normalized solution of (5), is excluded. We are therefore dealing with Case I; i.e., the integral equation (1) possesses a solution for every function $f(s)$. This solution is unique because a nonvanishing difference of two solutions would be a nontrivial solution of (5), in contradiction to the hypothesis that $r = 0$, i.e. that no such solution exists. The first Fredholm theorem is thus proved.

Now consider $r > 0$. Let $\psi_1, \psi_2, \cdots, \psi_r$ be r mutually orthogonal normalized solutions of (5). Then, since $A_n \rightrightarrows K$,[1] the functions

$$\delta_{ni}(s) = \psi_i(s) - \lambda \int A_n(s, t)\psi_i(t)\, dt$$

$$(i = 1, 2, \cdots, r; \qquad n = 1, 2, 3, \cdots)$$

satisfy $\delta_{ni}(s) \rightrightarrows 0$ for $n \to \infty$.

We now define $A'_n(s, t)$ by

$$A'_n(s, t) = A_n(s, t) + \frac{1}{\lambda} \sum_{i=1}^{r} \delta_{ni}(s)\psi_i(t);$$

then the $A'_n(s, t)$ are degenerate kernels which approximate the kernel $K(s, t)$ uniformly. It is easily seen that all these kernels $A'_n(s, t)$ possess the r null solutions $\psi_i(s)$.

There cannot be more than r linearly independent null solutions for sufficiently large n. For, if $\psi_{r+1,n}(s)$ were a sequence of additional null solutions, which may be assumed to be orthonormal to $\psi_1, \psi_2, \cdots, \psi_r$, then, by our convergence principle, there would exist a null solution of (5) orthogonal to $\psi_1, \psi_2, \cdots, \psi_r$ and therefore linearly independent of these functions; this would contradict the hypothesis that (5) has exactly r linearly independent null solutions.

Because of the validity of the Fredholm theorems for degenerate kernels the transposed homogeneous integral equations

$$\chi(s) = \lambda \int A'_n(t, s)\chi(t)\, dt \tag{19}$$

also possess exactly r linearly independent null solutions $\chi_{i,n}(s)$ $(i = 1, 2, \cdots, r)$ for sufficiently large n. These solutions may be

[1] The arrow $\to$ is occasionally used to denote convergence. We employ the double arrow $\rightrightarrows$ to denote uniform convergence.

taken to be orthonormal. Since the degenerate kernels $A'_n(t, s)$ converge uniformly to the kernel $K(t, s)$ we also obtain r mutually orthogonal null solutions $\chi_1(s), \chi_2(s), \cdots, \chi_r(s)$ for this kernel if we pass to the limit, employ our convergence principle, and utilize the equicontinuous and uniformly bounded set of functions $\chi_{i,n}(s)$. The transposed integral equation

$$\chi(s) = \lambda \int K(t, s)\chi(t)\, dt \tag{20}$$

cannot have more than r independent solutions, since otherwise we could, reversing the argument, show that (5) would necessarily also have more than r solutions.

Finally, we note that the conditions

$$(f, \chi_i) = \int f(s)\chi_i(s)\, ds = 0 \qquad (i = 1, 2, \cdots, r) \tag{21}$$

are certainly necessary for the solvability of the integral equation (1) for $r > 0$. This follows immediately if we multiply (1) by $\chi_i(s)$, integrate, and reverse the order of integration, taking (20) into account. To see that conditions (21) are sufficient we confine ourselves to those indices n for which $\lim_{n\to\infty} \chi_{i,n}(s) = \chi_i(s)$ $(i = 1, 2, \cdots, r)$. The $\epsilon_{in} = (f, \chi_{i,n})$ converge to zero with increasing n because of (21.) We now construct the functions

$$f_n(s) = f(s) - \sum_{i=1}^{r} \epsilon_{in}\, \chi_{i,n}(s);$$

they satisfy $(f_n, \chi_{i,n}) = 0 \quad (i = 1, 2, \cdots, r)$. Therefore the integral equation

$$f_n(s) = \varphi(s) - \lambda \int A'_n(s, t)\varphi(t)\, dt \tag{22}$$

possesses a solution $\rho_n(s)$ orthogonal to $\psi_1(s), \psi_2(s), \cdots, \psi_r(s)$ the Fredholm theorems for degenerate kernels are valid. We see that (ρ_n, ρ_n) is bounded independently of n; otherwise, the argument for Case II(a) would yield a solution of (5) orthogonal to $\psi_1(s)$, $\psi_2(s), \cdots, \psi_r(s)$, which is impossible by hypothesis. Therefore our convergence principle again permits us to pass to the limit in the integral equation, and because $f_n(s) \Rightarrow f(s)$, we conclude that (1) has a solution. This completes the proof of all the Fredholm theorems for our kernel $K(s, t)$.

That there exist cases in which the homogeneous equation possesses nontrivial solutions will be seen in the next section.

§4. *Symmetric Kernels and Their Eigenvalues*

The theory of integral equations, like the theory of bilinear forms considered in Chapter I, can be developed in greater detail if the kernel $K(s, t)$ is symmetric, i.e. if it satisfies the relation

$$K(s, t) = K(t, s). \tag{23}$$

The integral equation is then identical with the transposed equation. We shall give a theory of symmetric integral equations which does not depend on the preceding section.

Our main problem is to find values of the parameter λ for which the homogeneous integral equation (5) possesses a nontrivial (normalized) solution. As mentioned before, these values $\lambda = \lambda_i$ and the corresponding solutions are known as the eigenvalues and eigenfunctions, respectively, of the kernel $K(s, t)$. We shall now prove the theorem, analogous to that of Ch. I, §3: *Every continuous symmetric kernel that does not vanish identically possesses eigenvalues and eigenfunctions; their number is denumerably infinite if and only if the kernel is nondegenerate. All eigenvalues of a real symmetric kernel are real.*

1. Existence of an Eigenvalue of a Symmetric Kernel. We begin by proving that at least one eigenvalue exists. For this purpose we consider the "quadratic integral form"

$$J(\varphi, \varphi) = \iint K(s, t)\varphi(s)\varphi(t)\, ds\, dt \tag{24}$$

which takes the place of the quadratic form of Chapter I; φ is any function which is continuous or piecewise continuous in the basic domain. The Schwarz inequality yields the relation

$$[J(\varphi, \varphi)]^2 \leq (\varphi, \varphi)^2 \iint K^2(s, t)\, ds\, dt.$$

Therefore $J(\varphi, \varphi)$ is bounded in absolute value if we impose the condition

$$(\varphi, \varphi) = 1. \tag{25}$$

The integral form is equal to zero for all admissible functions φ if and only if the kernel itself vanishes identically. To see this we introduce the "bilinear integral form"

$$J(\varphi, \psi) = J(\psi, \varphi) = \iint K(s, t)\varphi(s)\psi(t)\, ds\, dt \tag{26}$$

and note the relation

$$J(\varphi + \psi, \varphi + \psi) = J(\varphi, \varphi) + 2J(\varphi, \psi) + J(\psi, \psi); \tag{27}$$

then clearly, if the quadratic form vanishes identically the bilinear form also does so. Taking $\psi(t) = \int K(s, t)\varphi(s)\, ds$ in (26), we find that

$$\int \left(\int K(s, t)\varphi(s)\, ds \right)^2 dt = 0;$$

therefore $\int K(s, t)\varphi(s)\, ds = 0$ for arbitrary φ. If, for a particular value of t, we now take φ to be equal to $K(s, t)$ we obtain the desired identity $K(s, t) = 0$.

If a kernel has the property that $J(\varphi, \varphi)$ can assume only positive or only negative values (unless φ vanishes identically) it is said to be *positive definite* or *negative definite*; otherwise it is said to be *indefinite*.

Let us now suppose that $J(\varphi, \varphi)$ can take on positive values. We consider the problem of finding a normalized function φ for which $J(\varphi, \varphi)$ assumes the greatest possible value. Since $J(\varphi, \varphi)$ is bounded, there certainly exists a least upper bound $\kappa_1 = 1/\lambda_1$ for the values of $J(\varphi, \varphi)$; we shall now show that this positive least upper bound is actually attained for a suitable function $\varphi(s)$. We suppose that the kernel $K(s, t)$ is approximated uniformly by a sequence of degenerate symmetric kernels

$$A_n(s, t) = \sum_{i,k=1}^{q_n} c_{ik}^{(n)} \omega_i(s)\omega_k(t), \qquad c_{ik}^{(n)} = c_{ki}^{(n)}$$

of the form described at the end of §1. The maximum problem for the integral form $J_n(\varphi, \varphi) = \iint A_n(s, t)\varphi(s)\varphi(t)\, ds\, dt$ with the sub-

sidiary condition (25) turns out to be equivalent to the corresponding problem for a quadratic form in q_n variables. Indeed, setting

$$(\varphi, \omega_i) = x_i \qquad (i = 1, 2, \cdots, q_n),$$

we obtain

$$J_n(\varphi, \varphi) = \sum_{i,k=1}^{q_n} c_{ik}^{(n)} x_i x_k, \tag{28}$$

an expression for $J_n(\varphi, \varphi)$ as a quadratic form in $x_1, x_2, \cdots, x_{q_n}$ to be maximized subject to condition (25). Now the Bessel inequality of Ch. II, §1, 3, applied to $\varphi(s)$ and the orthogonal system of functions $\omega_1(s), \omega_2(s), \cdots, \omega_{q_n}(s)$, gives

$$(\varphi, \varphi) \geq \sum_{i=1}^{q_n} x_i^2;$$

thus the variables in the quadratic form (28) must satisfy the condition $\sum_{i=1}^{q_n} x_i^2 \leq 1$. Therefore the maximum of the form is attained when $\sum_{i=1}^{q_n} x_i^2 = 1$, since otherwise the value of $J_n(\varphi, \varphi)$ would be increased if we multiply by a suitable factor. We are thus confronted with the problem of transformation to principal axes which was treated in Ch. I, §3. The maximum of the form is accordingly attained for a set of values $x_1, x_2, \cdots, x_{q_n}$ for which the equations

$$\sum_{k=1}^{q_n} c_{ik}^{(n)} x_k = \kappa_{1n} x_i \qquad (i = 1, 2, \cdots, q_n) \tag{29}$$

hold. The factor of proportionality κ_{1n} is moreover equal to the maximum of $J_n(\varphi, \varphi)$. To verify this we multiply (29) by x_i and take the sum, noting that the right side becomes κ_{1n} (since $\sum x_i^2 = 1$), while the left side becomes $J_n(\varphi, \varphi)$. Let us now set

$$\varphi_n(s) = x_1\omega_1(s) + x_2\omega_2(s) + \cdots + x_{q_n}\omega_{q_n}(s),$$

where $x_1, x_2, \cdots, x_{q_n}$ is the maximizing set of values. Because of the orthogonality of the ω_ν and because $\sum_{i=1}^{q_n} x_i^2 = 1$, this function is normalized. Equations (29) imply relation

$$\varphi_n(s) = \frac{1}{\kappa_{1n}} \int A_n(s, t)\varphi_n(t)\, dt \tag{30}$$

and conversely. For, (30) follows from (29) if we multiply by $\omega_i(s)$ and sum, noting that $x_i = (\varphi_n, \omega_i)$; if on the other hand we multiply

(30) by $\omega_i(s)$ and integrate we obtain (29). Thus the function $\varphi_n(s)$ is an eigenfunction of $A_n(s, t)$ belonging to the eigenvalue $\mu_{1n} = 1/\kappa_{1n}$; i.e.,

$$\varphi_n(s) = \mu_{1n} \int A_n(s, t)\varphi_n(t)\, dt. \tag{31}$$

We now let n increase beyond all bounds. Then κ_{1n} must converge to a number κ_1 which is the corresponding positive upper bound of $J(\varphi, \varphi)$; for, from the relation

$$| K(s, t) - A_n(s, t) | < \epsilon$$

and the Schwarz inequality it follows that

$$[J(\varphi, \varphi) - J_n(\varphi, \varphi)]^2 \leq \epsilon^2(b - a)^2,$$

where a and b are the limits of integration, provided $(\varphi, \varphi) \leq 1$. Thus, for sufficiently large n, the range of $J_n(\varphi, \varphi)$ coincides arbitrarily closely with the range of $J(\varphi, \varphi)$; the same must therefore be true of the upper bounds of the two ranges. Consequently $\lim_{n\to\infty} \kappa_{1n} = \kappa_1$ exists, and thus all the κ_{1n} lie below a fixed bound. Therefore, in virtue of (31) and of the considerations of §1, the functions $\varphi_n(s)$ are uniformly bounded for all n and form an equicontinuous set. By our convergence theorem it is possible to select a subsequence $\varphi_{n_1}, \varphi_{n_2}, \cdots$ which converges uniformly to a limit function $\psi_1(s)$. Passing to the limit in equation (30) and in the relations $J_n(\varphi_n, \varphi_n) = \kappa_{1n}$ and $(\varphi_n, \varphi_n) = 1$, we obtain the equations

$$\kappa_1\psi_1(s) = \int K(s, t)\psi_1(t)\, dt, \qquad (\psi_1, \psi_1) = 1 \tag{32}$$

and

$$J(\psi_1, \psi_1) = \kappa_1. \tag{33}$$

Thus the function $\psi_1(s)$ solves the maximum problem for the form $J(\varphi, \varphi)$; it is an eigenfunction of the kernel $K(s, t)$. Since it was assumed that $J(\varphi, \varphi)$ can take on positive values, κ_1 cannot be equal to zero. For any arbitrary function ψ we therefore have the inequality

$$J(\psi, \psi) \leq \kappa_1(\psi, \psi) \tag{34}$$

as is seen at once by normalizing ψ. (This method of proof is due to Holmgren.)

2. The Totality of Eigenfunctions and Eigenvalues. To obtain the other eigenvalues and eigenfunctions we proceed as follows:

We consider the problem of maximizing the integral $J(\varphi, \varphi)$ if, in addition to the condition $(\varphi, \varphi) = 1$, the condition

$$(\varphi, \psi_1) = 0$$

is imposed. We assume that under these conditions $J(\varphi, \varphi)$ can still take on positive values. By the second subsidiary condition, the range of the integral form $J(\varphi, \varphi)$ is more restricted in the present maximum problem than in the first; hence the maximum $\kappa_2 = 1/\mu_2$ cannot be greater than the previous maximum κ_1; i.e., $\kappa_2 \leq \kappa_1$ and $\mu_2 \geq \mu_1$. Here as in the first maximum problem the existence of a solution could be proved by reduction to a quadratic form and passage to the limit. It is, however, more convenient to reduce the problem directly to the problem of determining the first eigenvalue of another kernel.

We form the symmetric function

$$K_{(1)}(s, t) = K(s, t) - \frac{\psi_1(s)\psi_1(t)}{\mu_1}. \tag{35}$$

According to the above result the maximum problem

$$J_{(1)}(\varphi, \varphi) = \iint K_{(1)}(s, t)\varphi(s)\varphi(t)\, ds\, dt = \text{max.} = \kappa_2 = 1/\mu_2 \tag{36}$$

with the condition $(\varphi, \varphi) = 1$ can be solved by a function $\psi_2(s)$ which satisfies the homogeneous integral equation

$$\psi_2(s) = \mu_2 \int K_{(1)}(s, t)\psi_2(t)\, dt. \tag{37}$$

It is assumed here that $J_{(1)}(\varphi, \varphi)$ still admits positive values, so that $\kappa_2 > 0$. We write equation (37) in the form

$$\psi_2(s) = \mu_2 \int K(s, t)\psi_2(t)\, dt - \mu_2 \frac{\psi_1(s)}{\mu_1} (\psi_2, \psi_1),$$

multiply by $\psi_1(s)$, integrate with respect to s, reverse the order of integration in the iterated integral, and note that $(\psi_1, \psi_1) = 1$. The right-hand side then vanishes, so that we have

$$(\psi_1, \psi_2) = 0; \tag{38}$$

i.e. the eigenfunction ψ_2 is orthogonal to the eigenfunction ψ_1. Therefore we also have

$$\int K(s, t)\psi_2(t)\, dt = \int K_{(1)}(s, t)\psi_2(t)\, dt; \tag{39}$$

thus $\psi_2(s)$ is an eigenfunction of $K(s, t)$, and μ_2 the corresponding eigenvalue:

$$\psi_2(s) = \mu_2 \int K(s, t)\psi_2(t)\, dt. \tag{40}$$

Because of the relation $(\psi_2, \psi_1) = 0$, the value κ_2 can also be regarded as the maximum of the integral form $J_{(1)}(\varphi, \varphi)$ subject to the condition $(\varphi, \psi_1) = 0$; but under this condition $J_{(1)}(\varphi, \varphi) = J(\varphi, \varphi)$. Therefore the function ψ_2 is the solution of the maximum problem formulated at the beginning of this section.

We can continue in the same way, constructing the kernel

$$\begin{aligned} K_{(2)}(s, t) &= K_{(1)}(s, t) - \frac{\psi_2(s)\psi_2(t)}{\mu_2} \\ &= K(s, t) - \frac{\psi_1(s)\psi_1(t)}{\mu_1} - \frac{\psi_2(s)\psi_2(t)}{\mu_2} \end{aligned} \tag{41}$$

and seeking the maximum of the integral form

$$J_{(2)}(\varphi, \varphi) = \iint K_{(2)}(s, t)\varphi(s)\varphi(t)\, ds\, dt, \tag{42}$$

provided this form can still assume positive values. As above we obtain a normalized solution $\psi_3(s)$, orthogonal to ψ_1 and ψ_2, and a maximal value $\kappa_3 = 1/\mu_3$, which satisfy the homogeneous integral equation

$$\psi_3(s) = \mu_3 \int K(s, t)\psi_3(t)\, dt.$$

This solution could also be obtained from the problem of maximizing the original integral form by a function which is orthogonal to ψ_1 and ψ_2. Just as above it is seen that $\mu_2 \leq \mu_3$.

This procedure may be continued indefinitely, provided that all the kernels $K_{(1)}$, $K_{(2)}$, $\cdots$ give rise to integral forms which can

take on positive values. If, on the other hand, there occurs in this sequence a first kernel

$$K_{(m)}(s, t) = K(s, t) - \sum_{i=1}^{m} \frac{\psi_i(s)\psi_i(t)}{\mu_i} \tag{43}$$

for which $J_{(m)}(\varphi, \varphi) \leq 0$ for all φ, then we break off the sequence with the eigenfunction $\psi_m(s)$ and the eigenvalue μ_m.

In any case we have the result: The least positive eigenvalue μ_1 of the kernel $K(s, t)$ is the reciprocal of the maximum value κ_1 which the integral form $J(\varphi, \varphi)$ assumes under the subsidiary condition $(\varphi, \varphi) = 1$. This maximum is attained when φ is the first eigenfunction ψ_1 of $K(s, t)$. The eigenvalues μ_h $(h = 2, 3, \cdots)$, ordered in an increasing sequence, are defined recursively as the reciprocals of the maxima κ_h which $J(\varphi, \varphi)$ attains under the subsidiary conditions

$$(\varphi, \varphi) = 1, \qquad (\varphi, \psi_\nu) = 0 \qquad (\nu = 1, 2, 3, \cdots, h - 1).$$

The maximum κ_h is attained for $\varphi = \psi_h$, the h-th eigenfunction.

The sequence of positive eigenvalues breaks off when and if, in this sequence of maximum problems, one occurs for which $J(\varphi, \varphi)$ cannot take on positive values.

In an analogous manner we can now obtain a sequence of negative eigenvalues and corresponding eigenfunctions μ_{-1}, μ_{-2}, $\cdots$ and $\psi_{-1}(s)$, $\psi_{-2}(s)$, $\cdots$, provided the form $J(\varphi, \varphi)$ can assume negative values. We need only consider the minimum problems corresponding to the above maximum problems. We thus arrive at an infinite or terminating sequence of negative nonincreasing eigenvalues

$$\mu_{-1} \geq \mu_{-2} \geq \mu_{-3} \geq \cdots \tag{44}$$

and associated mutually orthogonal eigenfunctions

$$\psi_{-1}(s), \psi_{-2}(s), \cdots.$$

The eigenfunctions $\psi_h(s)$ $(h > 0)$ are orthogonal to the eigenfunctions $\psi_{-k}(s)$ $(k > 0)$. This fact results from the two equations

$$\kappa_h \psi_h(s) = \int K(s, t)\psi_h(t)\, dt,$$

$$\kappa_{-k} \psi_{-k}(s) = \int K(s, t)\psi_{-k}(t)\, dt$$

if we multiply the first by $\psi_{-k}(s)$ and the second by $\psi_h(s)$, subtract one from the other, and integrate; noting that $K(s, t) = K(t, s)$, we obtain

$$(\kappa_h - \kappa_{-k})(\psi_h , \psi_{-k}) = 0,$$

which proves the orthogonality, since $\kappa_h \neq \kappa_{-k}$.

The continuation of this procedure leads to a sequence of eigenvalues, which may be alternately positive and negative. We order these eigenvalues according to their absolute values and denote them by $\lambda_1 , \lambda_2 , \cdots$; thus $| \lambda_1 | \leq | \lambda_2 | \leq \cdots$. We denote the corresponding eigenfunctions by $\varphi_1 , \varphi_2 , \cdots$; they form an orthonormal system of functions.

If the kernel $K(s, t)$ possesses only a finite number of eigenvalues $\lambda_1, \lambda_2, \cdots , \lambda_n$ it must be degenerate and can be represented in the form

$$K(s, t) = \sum_{i=1}^{n} \frac{\varphi_i(s)\varphi_i(t)}{\lambda_i}. \tag{45}$$

For, according to the considerations of page 123, the kernel

$$K(s, t) - \sum_{i=1}^{n} \frac{\varphi_i(s)\varphi_i(t)}{\lambda_i} = \bar{K}(s, t)$$

must vanish identically, since both the maximum and the minimum of the associated integral form

$$\bar{J}(\varphi, \varphi) = \iint \bar{K}\ (s, t)\varphi(s)\varphi(t)\ ds\ dt$$

are equal to zero. Thus *a kernel which possesses only a finite number of eigenvalues and eigenfunctions is degenerate. Conversely, a degenerate kernel possesses only a finite number of eigenvalues and eigenfunctions.* Indeed, as we have seen above, the problem of determining the eigenvalues of such a kernel is equivalent to the eigenvalue problem for a quadratic form, where only a finite number of eigenvalues occurs (cf. Chapter I).

According to §1, an eigenvalue is called *multiple* or *degenerate*, specifically *r-fold degenerate*, if it is associated with r but with no more than r linearly independent eigenfunctions (which may be taken to be orthogonal). Every eigenvalue can have only a finite multiplicity or degeneracy. This theorem, proved in §1, may also

be obtained in the following way: We apply the Bessel inequality to the orthogonal system $\varphi_1, \varphi_2, \cdots$, obtaining

$$\int [K(s,t)]^2\,dt \geq \sum_{i=1}^{\infty} \left(\int K(s,t)\varphi_i(t)\,dt\right)^2 \tag{46}$$

or

$$\int [K(s,t)]^2\,dt \geq \sum_{i=1}^{\infty} \frac{[\varphi_i(s)]^2}{\lambda_i^2}. \tag{47}$$

Two conclusions follow: First, the series

$$T(s) = \sum_{i=1}^{\infty} \frac{[\varphi_i(s)]^2}{\lambda_i^2} \tag{48}$$

all of whose terms are positive, converges absolutely. Second, if we integrate with respect to s and note that $(\varphi_i, \varphi_i) = 1$ we find

$$\iint [K(s,t)]^2\,ds\,dt \geq \sum_{i=1}^{\infty} \frac{1}{\lambda_i^2}. \tag{49}$$

Thus *the sum of the reciprocals of the squares of the eigenvalues converges*. Therefore the eigenvalues can have no finite point of accumulation; if there are infinitely many eigenvalues, their absolute values must increase beyond all bounds and only a finite number of them can be equal to each other.

From this we shall conclude that the eigenvalues λ_i and eigenfunctions φ_i defined above by successive extremum problems constitute the totality of the real eigenvalues and eigenfunctions (it will be shown below that no complex eigenvalues can occur). Suppose that χ is an eigenfunction linearly independent of the φ_i, and that the corresponding eigenvalue is σ, which may be taken to be positive. Then, by the above argument, χ must be orthogonal to all eigenfunctions associated with the eigenvalues $\lambda_i \neq \sigma$. If, however, $\sigma = \mu_h$ is one of the eigenvalues defined above and is r-fold, that is if $\mu_{h-1} < \mu_h = \mu_{h+1} = \cdots = \mu_{h+r-1} < \mu_{h+r}$ ("degenerate case"), then χ may be replaced by an eigenfunction $\bar{\chi} = \chi + c_0\psi_h + \cdots + c_{r-1}\psi_{h+r-1}$ which is orthogonal to the eigenfunctions $\psi_h, \psi_{h+1}, \cdots, \psi_{h+r-1}$, since χ was assumed to be linearly independent of these functions. For simplicity we denote this function again by χ. Then in either of the cases considered χ is orthogonal to all eigen-

functions ψ_i. Therefore, for every n for which μ_{n+1} is positive, the relation

$$J(\chi, \chi) = \iint K(s, t)\chi(s)\chi(t)\, ds\, dt = \frac{1}{\sigma}(\chi, \chi) \leq \frac{1}{\mu_{n+1}}$$

holds, in virtue of the maximum property of the eigenvalues. Hence, if there are infinitely many positive eigenvalues μ_n, it follows from $\lim_{n\to\infty} \mu_n = \infty$ that $(\chi, \chi) = 0$, i.e. that χ vanishes identically. On the other hand, if there are only a finite number n of positive eigenvalues, then $J(\chi, \chi)$ cannot assume positive values under the subsidiary conditions $(\chi, \psi_i) = 0$ $(i = 1, 2, \cdots, n)$; we conclude once more that $(\chi, \chi) = 0$ and therefore that $\chi = 0$.

This proof holds equally well for negative σ. Therefore every eigenfunction which is orthogonal to all the ψ_i must vanish identically, which proves our assertion that the ψ_i represent all the possible eigenfunctions.

This fact has interesting consequences which will be useful later.

If $\eta_1(s)$, $\eta_2(s)$, $\cdots$ and $\zeta_1(s)$, $\zeta_2(s)$, $\cdots$ are two sequences of continuous (or piecewise continuous) functions whose norms lie below a fixed bound M, then the relation

$$\lim_{n\to\infty} J'_{(n)}(\eta_n, \zeta_n) = \lim_{n\to\infty} \iint K'_{(n)}(s, t)\eta_n(s)\zeta_n(t)\, ds\, dt = 0 \tag{50}$$

with the kernel

$$K'_{(n)}(s, t) = K(s, t) - \sum_{i=1}^{n} \frac{\varphi_i(s)\varphi_i(t)}{\lambda_i}$$

holds uniformly; i.e., for a given M, the smallness of the left side depends only on n.

In fact, by the maximum property of the eigenvalues and eigenfunctions, we have

$$|J'_{(n)}(\eta_n + \zeta_n, \eta_n + \zeta_n)| \leq \frac{1}{|\lambda_{n+1}|} N(\eta_n + \zeta_n) \leq \frac{4M}{|\lambda_{n+1}|}\,[1]$$

$$|J'_{(n)}(\eta_n, \eta_n)| \leq \frac{M}{|\lambda_{n+1}|}, \qquad |J'_{(n)}(\zeta_n, \zeta_n)| \leq \frac{M}{|\lambda_{n+1}|},$$

[1] It follows immediately from Schwarz's inequality that

$$N(\eta_n + \zeta_n) = (\eta_n, \eta_n) + (\zeta_n, \zeta_n) + 2(\eta_n, \zeta_n) \leq 4M.$$

and therefore, since $|\lambda_n| \to \infty$ and since

$$J'_{(n)}(\eta + \zeta, \eta + \zeta) = J'_{(n)}(\eta, \eta) + J'_{(n)}(\zeta, \zeta) + 2J'_{(n)}(\eta, \zeta),$$

the assertion follows at once.

Furthermore, *a kernel is positive definite if and only if all its eigenvalues are positive.* For, in this case and only in this case is the minimum of the integral form $J(\varphi, \varphi)$ for normalized φ positive. Thus $J(\varphi, \varphi)$ cannot have negative values.

Finally, *all eigenvalues of a real symmetric kernel are real.* Although this fact will become obvious in §5, we shall prove the theorem here in another, more direct way. The theorem states that there is no complex number $\lambda = p + iq$ where $q \neq 0$, with an associated complex function $\varphi(s) = \psi(s) + i\chi(s)$ (ψ and χ are real functions at least one of which does not vanish identically) such that $\varphi(s) = \lambda \int K(s, t)\varphi(t)\, dt$. For, the relation $\bar{\varphi}(s) = \bar{\lambda} \int K(s, t)\bar{\varphi}(t)\, dt$ would then have to hold for the conjugate complex quantities $\bar{\varphi}$ and $\bar{\lambda}$ and, as above,

$$0 = (\lambda - \bar{\lambda}) \int \varphi(s)\bar{\varphi}(s)\, ds = 2iq \int (\psi^2 + \chi^2)\, ds;$$

i.e., $q = 0$ which proves that λ is real.

3. Maximum-Minimum Property of Eigenvalues. As in the case of quadratic forms (Chapter I) each eigenvalue λ_n of a symmetric kernel and its associated eigenfunction can be directly characterized by a maximum-minimum problem without reference to preceding eigenvalues and eigenfunctions.

Let us consider first the positive eigenvalues μ_n of the kernel $K(s, t)$ and assume that there are at least n of these. We pose the problem to make $J(\varphi, \varphi)$ a maximum if φ is subject to the requirement $(\varphi, \varphi) = 1$, and, in addition, to the $n - 1$ restrictions

$$(\varphi, v_i) = 0 \qquad (i = 1, 2, \cdots, n - 1), \tag{51}$$

where $v_1, v_2, \cdots, v_{n-1}$ are any arbitrary given continuous functions. Certainly there exists an upper bound of $J(\varphi, \varphi)$ for some admissible function φ. This upper bound depends somehow on the choice of the functions $v_1, v_2, \cdots, v_{n-1}$; we may therefore denote it by $\kappa_n\{v_1, v_2, \cdots, v_{n-1}\}$ or briefly by $\kappa_n\{v_i\}$. In particular, for $v_i = \psi_i$, we have $\kappa_n\{v_i\} = \kappa_n$ by the theorems of the previous section; this

upper bound is attained for $\varphi = \psi_n(s)$. We now assert that, for any set of functions $v_1, v_2, \cdots, v_{n-1}$,

$$\kappa_n\{v_i\} \geq \kappa_n .$$

To prove this we construct an admissible function $\varphi(s) = c_1\psi_1(s) + c_2\psi_2(s) + \cdots + c_n\psi_n(s)$ as a linear combination of the eigenfunctions $\psi_1, \psi_2, \cdots, \psi_n$. The conditions $(\varphi, \varphi) = 1$ and (51) then take the form

$$\sum_{i=1}^{n} c_i^2 = 1, \qquad \sum_{i=1}^{n} c_i(\psi_i, v_h) = 0 \qquad (h = 1, 2, \cdots, n-1).$$

This is a system of $n - 1$ homogeneous linear equations in n unknowns c_i with a normality condition; such a system always has a solution. If the resulting function φ is substituted in $J(\varphi, \varphi)$, one obtains

$$J(\varphi, \varphi) = \sum_{i,k=1}^{n} c_i c_k J(\psi_i, \psi_k).$$

Since $J(\psi_i, \psi_i) = 1/\mu_i$ and $J(\psi_i, \psi_k) = 0$ for $i \neq k$ it follows that

$$J(\varphi, \varphi) = \sum_{i=1}^{n} \frac{c_i^2}{\mu_i} = \sum_{i=1}^{n} c_i^2 \kappa_i \geq \kappa_n \sum_{i=1}^{n} c_i^2 = \kappa_n .$$

The maximum of $J(\varphi, \varphi)$ is thus certainly at least equal to κ_n, and we obtain the result: *The n-th positive eigenvalue* λ_n *of* $K(s, t)$ *is the reciprocal of the least value* κ_n *which* $\kappa_n\{v_i\}$ *attains as the functions* v_i *are varied; here* $\kappa_n\{v_i\}$ *is defined as the maximum value (or least upper bound) of* $J(\varphi, \varphi)$ *for normed functions* $\varphi(s)$ *subject to the* $n - 1$ *further conditions* (51). *The minimum of this maximum is attained for*

$$v_1 = \psi_1, \qquad v_2 = \psi_2, \quad \cdots, \qquad v_{n-1} = \psi_{n-1} \text{ and } \varphi = \psi_n .$$

In the same way we define the negative eigenvalues and the associated eigenfunctions ψ_{-n} $(n > 0)$ in terms of the maximum of the minimum of $J(\varphi, \varphi)$, subject to the corresponding conditions.

Among the many applications of the above result we mention the following theorem which can be deduced from the maximum-minimum property of the eigenvalues: *If to a kernel* $K(s, t)$ *a positive definite kernel* $K^+(s, t)$ *(or a negative definite kernel* $K^-(s, t)$*) is added, then each positive (or negative) reciprocal eigenvalue of the sum* $K + K^+$ *(or* $K + K^-$*) is not less (or not greater) than the corresponding reciprocal*

eigenvalue of the kernel K.[1] The proof is analogous to that of the corresponding theorem in Ch. 1, §4.

§5. *The Expansion Theorem and Its Applications*

1. Expansion Theorem. If we knew that in analogy with the transformation to principal axes of a quadratic form, the kernel could be expanded in a series

$$K(s, t) = \sum_{i=1}^{\infty} \frac{\varphi_i(s)\varphi_i(t)}{\lambda_i} \tag{52}$$

which would converge uniformly in each variable, then it would follow that every function $g(s)$ of the form

$$g(s) = \int K(s, t)h(t)\, dt, \tag{53}$$

where $h(t)$ is any piecewise continuous function, could be expanded in the series

$$g(s) = \sum_{i=1}^{\infty} g_i \varphi_i(s), \qquad g_i = (g, \varphi_i) = \frac{(h, \varphi_i)}{\lambda_i}.$$

However the relation (52) is not valid in general; we are thus forced to adopt a somewhat different approach to justify the expansion for $g(s)$. Let $g(s)$ be an integral transform of $h(t)$ with symmetric kernel $K(s, t)$ as defined in equation (53); let $h_i = (h, \varphi_i)$ be the expansion coefficients of h with respect to the orthonormal eigenfunctions $\varphi_1, \varphi_2, \cdots$, and let

$$g_i = (g, \varphi_i) = \frac{h_i}{\lambda_i}$$

be the expansion coefficients of g; in accordance with the Bessel inequality, the series $\sum_{i=1}^{\infty} h_i^2$ converges. By equation (47), the sum

$$T(s) = \sum_{i=1}^{\infty} \frac{[\varphi_i(s)]^2}{\lambda_i^2}$$

[1] See H. Weyl, Das asymptotische Verteilungsgesetz der Eigenwerte linearer partieller Differentialgleichungen (mit einer Anwendung auf die Theorie der Hohlraumstrahlung), Math. Ann., Vol. 71, 1912, pp. 441–479.

converges and is uniformly bounded in s (by $\max_{a\le s\le b} \int [K(s, t)]^2\, dt$); in view of the Schwarz inequality, we have

$$\left[\frac{h_n \varphi_n(s)}{\lambda_n} + \cdots + \frac{h_m \varphi_m(s)}{\lambda_m}\right]^2 \le (h_n^2 + \cdots h_m^2)\left(\frac{[\varphi_n(s)]^2}{\lambda_n^2} + \cdots + \frac{[\varphi_m(s)]^2}{\lambda_m^2}\right).$$

Since the remainder $h_n^2 + h_{n+1}^2 + \cdots + h_m^2$ is arbitrarily small provided n is sufficiently large, and since $[\varphi_n(s)]^2/\lambda_n^2 + \cdots + [\varphi_m(s)]^2/\lambda_m^2$ is always below a bound independent of s, it follows that the series

$$\sum_{i=1}^{\infty} g_i \varphi_i(s) = \sum_{i=1}^{\infty} (h_i/\lambda_i)\varphi_i(s)$$

converges absolutely and uniformly. Its sum

$$\gamma(s) = \lim_{n\to\infty} \sum_{i=1}^{n} g_i \varphi_i(s) = \lim_{n\to\infty} \gamma_n(s)$$

is a continuous function of s. It remains to be shown that $\gamma(s)$ is identical with $g(s)$. For this purpose we construct the kernel

$$K_{(n)}(s, t) = K(s, t) - \sum_{i=1}^{n} \frac{\varphi_i(s)\varphi_i(t)}{\lambda_i}\,^{1}$$

so that

$$g(s) - \gamma_n(s) = \int K_{(n)}(s, t)h(t)\, dt;$$

we multiply this equation by an arbitrary continuous function $w(s)$ and integrate with respect to s. Because of relation (50) the right-hand side in the relation

$$\int w(s)\big(g(s) - \gamma_n(s)\big)\, ds = \iint K_{(n)}(s, t)h(t)w(s)\, ds\, dt$$

converges to zero and we obtain

$$\int w(s)\big(g(s) - \gamma(s)\big)\, ds = 0$$

[1] We shall henceforth use this notation for the functions which were denoted by $K'_{(n)}$ on p. 131.

since $\gamma_n(s) \Rightarrow \gamma(s)$. This equation holds for an arbitrary function $w(s)$, in particular for $w(s) = g(s) - \gamma(s)$. Since $g(s) - \gamma(s)$ is continuous the relation $(g - \gamma, g - \gamma) = 0$ can hold only if $g(s) - \gamma(s)$ vanishes identically, as was to be proved. We have thus obtained the fundamental expansion theorem:

Every continuous function $g(s)$ which, as in (53), is an integral transform with symmetric kernel $K(s, t)$ of a piecewise continuous function $h(t)$, can be expanded in a series in the eigenfunctions of $K(s, t)$; this series converges uniformly and absolutely.

2. Solution of the Inhomogeneous Linear Integral Equation. As an application of this theorem we shall derive the formula for the solution of the inhomogeneous integral equation (1),

$$f(s) = \varphi(s) - \lambda \int K(s, t)\varphi(t)\, dt.$$

We assume initially that the parameter λ is not equal to any of the eigenvalues λ_i. If the continuous function φ with the expansion coefficients (φ, φ_i) were a solution of the integral equation, then, by the expansion theorem applied to $h(t) = \lambda\varphi(t)$, the function $\varphi(s) - f(s) = g(s)$ would be given by the uniformly and absolutely convergent series

$$(54) \qquad g(s) = \varphi(s) - f(s) = \sum_{i=1}^{\infty} c_i\varphi_i(s) = \lambda \int K(s, t)\varphi(t)\, dt,$$

where $c_i = (g, \varphi_i)$. But, because of (54), we must have

$$c_i = (g, \varphi_i) = \lambda \iint K(s, t)\varphi_i(s)\varphi(t)\, ds\, dt$$

$$= \frac{\lambda}{\lambda_i}(\varphi_i, \varphi) = \frac{\lambda}{\lambda_i}(\varphi_i, f) + \frac{\lambda}{\lambda_i}(\varphi_i, g),$$

from which it follows that

$$(55) \qquad c_i = f_i \frac{\lambda}{\lambda_i - \lambda} \qquad (f_i = (\varphi_i, f)).$$

We thus obtain the series expansion for φ,

$$(56) \qquad \varphi(s) = f(s) + \lambda \sum_{i=1}^{\infty} \frac{f_i}{\lambda_i - \lambda} \varphi_i(s),$$

which must represent the solution of (1). That this really is so, i.e. that (56) actually provides the solution of (1), is easily seen:

The series converges absolutely and uniformly. To prove this we need only note that, for sufficiently large i and a given arbitrary λ, the relation $|\lambda_i - \lambda| > |\lambda_i|/2$ is sure to hold. Thus, except for a finite number of terms, the series $2|\lambda| \sum_{i=1}^{\infty} |f_i||\varphi_i(s)|/|\lambda_i|$ is a majorant of our series; the uniform convergence of this majorant has already been proved. If we now substitute the series (56) in (1), we verify immediately that the equation (1) is satisfied.

In conformity with the theory of §3 this solution fails only if $\lambda = \lambda_i$ is an eigenvalue; it remains valid even in this case if $f(s)$ is orthogonal to all eigenfunctions φ_i belonging to the eigenvalue λ_i.

Since, by §3, the integral equation (1) fails to have solutions for certain functions $f(s)$ only if λ is an eigenvalue, it follows that there can be no eigenvalues of our kernel other than the values λ_i. The assertion that all the eigenvalues of a real symmetric kernel are real, which was proved on page 132, has now become self-evident.

3. Bilinear Formula for Iterated Kernels. Another application of the expansion theorem is obtained setting $h(\sigma) = K(\sigma, t)$. For the "iterated kernel"

$$K^{(2)}(s, t) = \int K(s, \sigma)K(\sigma, t)\, d\sigma$$

we then have the expansion

$$K^{(2)}(s, t) = \sum_{i=1}^{\infty} \frac{\varphi_i(s)}{\lambda_i} \int K(\sigma, t)\varphi_i(\sigma)\, d\sigma$$

or

$$K^{(2)}(s, t) = \sum_{i=1}^{\infty} \frac{\varphi_i(s)\varphi_i(t)}{\lambda_i^2}. \tag{57}$$

In the same way, the subsequent iterated kernels

$$K^{(3)}(s, t) = \int K^{(2)}(s, \sigma)K(\sigma, t)\, d\sigma$$

$$= \iint K(s, \sigma_1)K(\sigma_1, \sigma_2)K(\sigma_2, t)\, d\sigma_1\, d\sigma_2,$$

. . . . . .

$$K^{(n)}(s, t) = \int K^{(n-1)}(s, \sigma)K(\sigma, t)\, d\sigma$$

$$= \int \cdots \int K(s, \sigma_1)K(\sigma_1, \sigma_2) \cdots K(\sigma_{n-1}, t)\, d\sigma_1 \cdots d\sigma_{n-1}$$

admit the expansions

$$K^{(n)}(s, t) = \sum_{i=1}^{\infty} \frac{\varphi_i(s)\varphi_i(t)}{\lambda_i^n} \qquad (n = 2, 3, \cdots), \tag{58}$$

all of which converge absolutely and uniformly both in s and in t, and uniformly in s and t together (see subsection 4).

From (57) it follows that

$$K^{(2)}(s, s) = \sum_{i=1}^{\infty} \frac{[\varphi_i(s)]^2}{\lambda_i^2};$$

thus

$$\lim \left(K^{(2)}(s, s) - \sum_{i=1}^{n} \frac{[\varphi_i(s)]^2}{\lambda_i^2} \right) = 0.$$

But this means that

$$\lim_{n\to\infty} \int \left[K(s, t) - \sum_{i=1}^{n} \frac{\varphi_i(s)\varphi_i(t)}{\lambda_i} \right]^2 dt = 0; \tag{59}$$

i.e., the series $\sum_{i=1}^{\infty} \varphi_i(s)\varphi_i(t)/\lambda_i$ converges in the mean to $K(s, t)$. If, for fixed s, this series converges uniformly in t and therefore represents a continuous function $L(s, t)$ of t, we must have $K = L$. For, in this case we can carry out the passage to the limit in (59) under the integral sign, obtaining $\int [K(s, t) - L(s, t)]^2 \, dt = 0$ and, therefore, $K - L = 0$.

4. Mercer's Theorem.[1] Formula (58) for $n > 1$ may be regarded as a substitute for equation (52), which in general cannot be proved. However, for an important special case, we can state the following theorem: *If $K(s, t)$ is a definite continuous symmetric kernel, or if it has only a finite number of eigenvalues of one sign, the expansion* (52) *is valid and converges absolutely and uniformly.*

In the proof we shall assume initially that $K(s, t)$ is positive definite, i.e. that all eigenvalues λ_i are positive. We first remark that, for every positive definite kernel $H(s, t)$, the inequality $H(s, s) \geq 0$ holds. For, if $H(s_0, s_0) < 0$, there would be a neighborhood of the point $s = s_0$, $t = s_0$, say $|s - s_0| < \epsilon$, $|t - s_0| < \epsilon$, such that $H(s, t) < 0$ everywhere in this region. We then define the function

[1] T. Mercer, Functions of positive and negative type and their connection with the theory of integral equations, Trans. Lond. Phil. Soc. (A), Vol. 209, 1909, pp. 415–446.

$\varphi(s)$ by $\varphi(s) = 1$ for $|s - s_0| \leq \epsilon$, and by $\varphi(s) = 0$ elsewhere. For this function we have

$$\iint H(s, t)\varphi(s)\varphi(t)\, ds\, dt < 0$$

in contradiction to the hypothesis that H is positive definite. If we now apply this result to the positive definite kernel

$$H = K(s, t) - \sum_{i=1}^{n} \frac{\varphi_i(s)\varphi_i(t)}{\lambda_i},$$

we obtain

$$K(s, s) - \sum_{i=1}^{n} \frac{[\varphi_i(s)]^2}{\lambda_i} \geq 0.$$

Therefore the series

$$\sum_{i=1}^{n} \frac{[\varphi_i(s)]^2}{\lambda_i},$$

all the terms of which are positive, converges for every value of s. Because of the relation

$$\left(\frac{\varphi_n(s)}{\sqrt{\lambda_n}}\frac{\varphi_n(t)}{\sqrt{\lambda_n}} + \cdots + \frac{\varphi_m(s)}{\sqrt{\lambda_m}}\frac{\varphi_m(t)}{\sqrt{\lambda_m}}\right)^2 \leq \left(\frac{[\varphi_n(s)]^2}{\lambda_n} + \cdots + \frac{[\varphi_m(s)]^2}{\lambda_m}\right)\left(\frac{[\varphi_n(t)]^2}{\lambda_n} + \cdots + \frac{[\varphi_m(t)]^2}{\lambda_m}\right)$$

(Schwarz inequality) the series $\sum_{i=1}^{\infty} \varphi_i(s)\varphi_i(t)/\lambda_i$ also converges absolutely; it converges uniformly in t for fixed s and in s for fixed t. Thus the function defined by this series is continuous in s for fixed t and conversely. In view of subsection 3, it is therefore equal to the kernel K.

Finally we show that this series also converges uniformly in both variables together; because of the above inequality it suffices to verify the uniformity of the convergence of the series

$$\sum_{i=1}^{\infty} \frac{[\varphi_i(s)]^2}{\lambda_i}.$$

But, according to what we have just proved, this series is equal to $K(s, s)$, which is a continuous function. Now the following theorem

holds:[1] if a series of positive continuous functions of one variable converges to a continuous function, the convergence of the series is uniform in the interval in question. This theorem leads immediately to the result stated.

The existence of a finite number of negative eigenvalues cannot alter the convergence of the series (52), since the kernel can be made positive definite by subtraction of the terms $\varphi_i(s)\varphi_i(t)/\lambda_i$ belonging to negative eigenvalues. Thus our convergence theorem is completely proved.

§6. *Neumann Series and the Reciprocal Kernel*

The preceding theory of integral equations implies a prescription for actually computing solutions with arbitrary accuracy (see also §8). It does not give the solutions, however, in an elegant closed form such as was obtained in the theory of equations in Chapter I. To find comparable explicit solutions we shall use a method analogous to that of Chapter I. We rewrite the integral equation (1), inserting in place of $\varphi(t)$ in the integral the expression obtained from (1), and repeat this procedure indefinitely. With the aid of the iterated kernels we thus write (1) in the form

$$\begin{aligned}\varphi(s) &= f(s) + \lambda \int K(s, t)f(t)\,dt + \lambda^2 \int K^{(2)}(s, t)\varphi(t)\,dt \\ &= f(s) + \lambda \int K(s, t)f(t)\,dt + \lambda^2 \int K^{(2)}(s, t)f(t)\,dt \\ &\qquad + \lambda^3 \int K^{(3)}(s, t)\varphi(t)\,dt \\ &= \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\end{aligned}$$

and see, just as in Chapter I, that the solution is given by the infinite series

$$(60)\quad \varphi(s) = f(s) + \lambda \int K(s, t)f(t)\,dt + \lambda^2 \int K^{(2)}(s, t)f(t)\,dt + \cdots,$$

provided that this series converges uniformly. If, moreover, we postulate the uniform convergence of

$$(61)\quad \mathsf{K}(s, t) = K(s, t) + \lambda K^{(2)}(s, t) + \lambda^2 K^{(3)}(s, t) + \cdots,$$

[1] Compare footnote on p. 57.

then the solution of the integral equation (1)

$$f(s) = \varphi(s) - \lambda \int K(s, t)\varphi(t)\, dt$$

is represented by the "*reciprocal integral equation*"

$$\varphi(s) = f(s) + \lambda \int \mathsf{K}(s, t)f(t)\, dt. \tag{62}$$

The function $\mathsf{K}(s, t) = \mathsf{K}(s, t; \lambda)$ is therefore called the *reciprocal* or *resolvent* kernel.

Series (60) and (61) are known as *Neumann's series*. They converge uniformly if $|\lambda|$ is sufficiently small, e.g. if $|\lambda| < 1/(b - a)M$, where M is an upper bound of the absolute value of $K(s, t)$. Thus, for sufficiently small $|\lambda|$, the reciprocal kernel is an analytic function of λ. It satisfies the relations

$$\mathsf{K}(s, t; \lambda) = K(s, t) + \lambda \int K(s, \sigma)\mathsf{K}(\sigma, t; \lambda)\, d\sigma,$$

$$\mathsf{K}(s, t; \lambda) = K(s, t) + \lambda \int K(\sigma, t)\mathsf{K}(s, \sigma; \lambda)\, d\sigma, \tag{63}$$

$$\mathsf{K}(s, t; \lambda) - \mathsf{K}(s, t; \lambda') = (\lambda - \lambda') \int \mathsf{K}(s, \sigma; \lambda)\mathsf{K}(\sigma, t; \lambda')\, d\sigma;$$

this can be immediately verified by substitution.

If the kernel $K(s, t)$ is symmetric we can find a remarkable form for the resolvent kernel, which shows how the analytic function K depends on λ. Assuming $|\lambda|$ to be sufficiently small, we make use of the series expansion (58) for the symmetric kernels $K^{(2)}(s, t)$, $K^{(3)}(s, t)$, $\cdots$ and take the sum of the geometric series occurring in (61); we immediately obtain

$$\mathsf{K}(s, t; \lambda) = K(s, t) + \lambda \sum_{i=1}^{\infty} \frac{\varphi_i(s)\varphi_i(t)}{\lambda_i(\lambda_i - \lambda)}. \tag{64}$$

By proceeding as in §5, 1 and 2, we see that the series on the right converges for every value of λ which is not an eigenvalue, and the convergence is uniform in s and t.

Relation (64) which has so far been proved only for sufficiently small $|\lambda|$ provides the analytic continuation of the resolvent $\mathsf{K}(s, t; \lambda)$ into the entire complex λ-plane, with the eigenvalues λ_i

appearing as simple poles. Thus (64) represents the decomposition of the resolvent into partial fractions, and we may express the result as follows: *The resolvent of a symmetric kernel is a meromorphic function of* λ *which possesses simple poles at the eigenvalues of the integral equation.* Its residues at the poles λ_i provide the eigenfunctions belonging to these eigenvalues. From the Neumann series and the representation (64) it follows that the radius of convergence of the Neumann series is equal to the absolute value of the eigenvalue with the smallest square.

According to the general theory of functions, the resolvent $K(s, t; \lambda)$, being a meromorphic function, can be represented as the quotient of two integral transcendental functions; it is to be expected that each of these functions can be expanded in an everywhere convergent power series the coefficients of which can be obtained directly from the given kernel. In the algebraic case we have such a representation in the formulas of Ch. I, §2. It is reasonable to suppose that very similar formulas can be obtained here. We may expect, furthermore, that these formulas are in no way restricted to symmetric kernels only, but remain valid for arbitrary continuous unsymmetric kernels. Such representations have been obtained by Fredholm, who then used them as the starting point of the theory. In the next section we shall derive Fredholm's formulas, again approximating the kernel uniformly by degenerate kernels $A_n(s, t)$ and then passing to the limit $n \to \infty$.[1]

§7. *The Fredholm Formulas*

Since we shall not make any use of the Fredholm formulas later on, we omit some intermediate calculations involving determinants.[2]

We shall employ essentially the same procedure and terminology as in Ch. I, §2. For a degenerate kernel

$$K(s, t) = A(s, t) = \sum_{p=1}^{n} \alpha_p(s)\beta_p(t),$$

[1] This method was first employed by Goursat: Sur un cas élémentaire de l'équation de Fredholm, Bull. Soc. math. France, Vol. 35, 1907, pp. 163–173. Compare also H. Lebesgue, Sur la méthode de M. Goursat pour la résolution de l'équation de Fredholm, ibid., Vol. 36, 1909, pp. 3–19.

[2] See G. Kowalewski, Einführung in die Determinantentheorie, Veit Leipzig, 1909.

the integral equation (1) goes over into

$$\varphi(s) = f(s) + \lambda \sum_{p=1}^{n} x_p \alpha_p(s) = f(s) + \lambda E(\alpha(s), x) \tag{65}$$

if we set $x_p = (\varphi, \beta_p)$ as before. Making use of the notation $y_p = (f, \beta_p)$, $k_{pq} = (\alpha_q, \beta_p)$, we obtain the system of equations

$$y_p = x_p - \lambda \sum_{q=1}^{n} k_{pq} x_q \tag{66}$$

for the quantities x_p. The solution of this system is given by

$$E(u, x) = -\frac{\Delta(u, y; \lambda)}{\Delta(\lambda)}$$

and, thus, the solution of (1) is

$$\varphi(s) = f(s) + \lambda E(\alpha(s), x) = f(s) - \lambda \frac{\Delta(\alpha(s), y; \lambda)}{\Delta(\lambda)}; \tag{67}$$

in (67), we have

$$\begin{aligned} &\Delta(u, y; \lambda) \\ &\quad = \Delta_1(u, y) - \lambda \Delta_2(u, y) + \cdots + (-1)^{n-1} \lambda^{n-1} \Delta_n(u, y), \\ &\Delta(\lambda) = 1 - \lambda \Delta_1 + \cdots + (-1)^n \lambda^n \Delta_n ; \end{aligned} \tag{68}$$

with

$$\begin{aligned} \Delta_h(u, y) &= \sum \begin{vmatrix} 0 & u_{p_1} & u_{p_2} & \cdots & u_{p_h} \\ y_{p_1} & k_{p_1 p_1} & k_{p_1 p_2} & \cdots & k_{p_1 p_h} \\ y_{p_2} & k_{p_2 p_1} & k_{p_2 p_2} & \cdots & k_{p_2 p_h} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ y_{p_h} & k_{p_h p_1} & k_{p_h p_2} & \cdots & k_{p_h p_h} \end{vmatrix}, \\ \Delta_h &= \sum \begin{vmatrix} k_{p_1 p_1} & k_{p_1 p_2} & \cdots & k_{p_1 p_h} \\ k_{p_2 p_1} & k_{p_2 p_2} & \cdots & k_{p_2 p_h} \\ \cdots & \cdots & \cdots & \cdots \\ k_{p_h p_1} & k_{p_h p_2} & \cdots & k_{p_h p_h} \end{vmatrix}. \end{aligned} \tag{69}$$

In (69) the summation indices $p_1, p_2, \cdots, p_h$ run independently from 1 to n with $p_1 < p_2 < \cdots < p_h$.

The sum of determinants $\Delta(\alpha(s), y; \lambda)$ may evidently be written in the form $\int \Delta[\alpha(s), \beta(t); \lambda] f(t)\, dt$, so that the solution (67) of the integral equation takes on the form

$$\varphi(s) = f(s) + \lambda \int K(s, t; \lambda) f(t)\, dt \tag{62'}$$

with the resolvent

$$K(s, t; \lambda) = -\frac{\Delta(\alpha(s), \beta(t); \lambda)}{\Delta(\lambda)} = \frac{D(s, t; \lambda)}{D(\lambda)}. \tag{70}$$

In the formulas (69) the summation, instead of extending only over all sets of indices for which $p_1 < p_2 < \cdots < p_h$, may just as well extend over all possible combinations, the result being divided by $h!$. Making use of this fact and of the definition of k_{pq}, and employing certain simple theorems concerning determinants, we obtain the formulas

$$\begin{aligned} D(s, t; \lambda) &= -\ \Delta(\alpha(s), \beta(t); \lambda) \\ &= D_0(s, t) - \frac{1}{1!} D_1(s, t)\lambda + \frac{1}{2!} D_2(s, t)\lambda^2 - \cdots \\ &\qquad + \frac{(-1)^{n-1}}{(n-1)!} D_{n-1}(s, t)\lambda^{n-1}, \\ D(\lambda) &= \Delta(\lambda) \\ &= 1 - \frac{1}{1!} D_1\lambda + \frac{1}{2!} D_2\lambda^2 - \cdots + \frac{(-1)^n}{n!} D_n\lambda^n, \end{aligned} \tag{71}$$

where we have used the abbreviations

$$D_h(s, t) = \iint \cdots \int \begin{vmatrix} A(s, t) & A(s, s_1) & \cdots & A(s, s_h) \\ A(s_1, t) & A(s_1, s_1) & \cdots & A(s_1, s_h) \\ \cdots & \cdots & \cdots & \cdots \\ A(s_h, t) & A(s_h, s_1) & \cdots & A(s_h, s_h) \end{vmatrix} ds_1 ds_2 \cdots ds_h$$

$$D_h = \iint \cdots \int \begin{vmatrix} A(s_1, s_1) & A(s_1, s_2) & \cdots & A(s_1, s_h) \\ A(s_2, s_1) & A(s_2, s_2) & \cdots & A(s_2, s_h) \\ \cdots & \cdots & \cdots & \cdots \\ A(s_h, s_1) & A(s_h, s_2) & \cdots & A(s_h, s_h) \end{vmatrix} ds_1 ds_2 \cdots ds_h \tag{72}$$

for $h = 1, 2, \cdots, n$, and $D_0(s, t) = A(s, t)$.

Thus the integral rational functions $D(s, t; \lambda)$ and $D(\lambda)$ of λ have been written down explicitly in terms of the given kernel. The series (71) may be formally continued as infinite series since the quantities D_h for $h > n$ and $D_h(s, t)$ for $h > n - 1$, as defined by (72), all vanish for the degenerate kernel $A(s, t) = \sum_{p=1}^{n} \alpha_p(s)\beta_p(t)$.

Now, if the arbitrary continuous kernel $K(s, t)$ is approximated uniformly by a sequence of degenerate kernels, then the corresponding expressions (72) converge to the corresponding determinants of the kernel $K(s, t)$. The infinite series

$$D(s, t; \lambda) = D_0(s, t) - \frac{\lambda}{1!} D_1(s, t) + \cdots,$$

(73)

$$D(\lambda) = 1 - \frac{\lambda}{1!} D_1 + \frac{\lambda^2}{2!} D_2 - \cdots,$$

where A is replaced by K in (72), represent integral transcendental functions of λ for the nondegenerate kernel $K(s, t)$. To see this it is only necessary to show that they converge for every value of λ. If $| K(s, t) | \leq M$ for all s and t, then, by Hadamard's inequality for determinants (Ch. I, §5, 2),

$$| D_h(s, t) | \leq \sqrt{(h + 1)^{h+1}}\, M^{h+1}(b - a)^h,$$

$$| D_h | \leq \sqrt{h^h}\, M^h (b - a)^h.$$

Now the series

$$\sum_{h=0}^{\infty} \sqrt{(h + 1)^{h+1}}\, M^{h+1}(b - a)^h \frac{\lambda^h}{h!}, \qquad 1 + \sum_{h=1}^{\infty} \sqrt{h^h}\, M^h (b - a)^h \frac{\lambda^h}{h!}$$

converge for every value of λ[1] and are majorants of the series of the absolute values of the terms of (73). Thus the assertion is proved. It follows that, for every value of λ, the convergence relations

$$\lim_{n \to \infty} D_n(s, t; \lambda) = D(s, t; \lambda), \qquad \lim_{n \to \infty} D_n(\lambda) = D(\lambda)$$

hold uniformly in s and t, where the quantities with index n refer to the n-th approximating degenerate kernel $A_n(s, t)$ and those without

[1] This can be seen from the fact that $1/h! < e^h/h^h$, since the term $h^h/h!$ occurs in the expansion of e^h. Therefore the h-th root of the coefficient of λ^h in the series on the right is less than $M(b - a)e/h^{1/2}$ and, consequently, converges to zero for $h \to \infty$; the same is true for the other series.

index to $K(s, t)$. Therefore, as long as λ is not one of the zeros λ_i of $D(\lambda)$, we have, for the resolvent of the kernel $K(s, t)$,

$$(74) \qquad \mathsf{K}(s, t; \lambda) = \frac{D_0(s, t) - \frac{\lambda}{1!} D_1(s, t) + \cdots}{1 - \frac{\lambda}{1!} D_1 + \frac{\lambda^2}{2!} D_2 - \cdots} = \lim_{n \to \infty} \mathsf{K}_n(s, t; \lambda);$$

thus, for the solution of an integral equation with arbitrary kernel $K(s, t)$ we obtain the formula

$$(75) \qquad \varphi(s) = f(s) + \lambda \int \mathsf{K}(s, t; \lambda) f(t)\, dt.$$

The above formulas are known as the *Fredholm formulas* in honor of their discoverer. Evidently the relation

$$(76) \qquad D_h = \int D_{h-1}(s, s)\, ds$$

holds. Note that we also have[1]

$$(77) \qquad D'(\lambda) = - \int D(s, s; \lambda)\, ds$$

and that the m-th derivative is given by

$$(78) \quad D^{(m)}(\lambda) = (-1)^m \iint \cdots \int D\left(\begin{matrix} s_1, s_2, \cdots, s_m \\ s_1, s_2, \cdots, s_m \end{matrix} \middle| \lambda\right) ds_1 ds_2 \cdots ds_m,$$

where

$$(79) \qquad D\left(\begin{matrix} s_1, s_2, \cdots, s_m \\ t_1, t_2, \cdots, t_m \end{matrix} \middle| \lambda\right) = \sum_{h=1}^{\infty} \frac{(-\lambda)^h}{h!} D_h\left(\begin{matrix} s_1, s_2, \cdots, s_m \\ t_1, t_2, \cdots, t_m \end{matrix}\right)$$

and

$$D_h\left(\begin{matrix} s_1, s_2, \cdots s_m \\ t_1, t_2, \cdots t_m \end{matrix}\right)$$

(80)

$$= \int\int \cdots \int \begin{vmatrix} K(s_1, t_1) \cdots K(s_1, t_m) & K(s_1, \sigma_1) \cdots K(s_1, \sigma_h) \\ K(s_2, t_1) \cdots K(s_2, t_m) & K(s_2, \sigma_1) \cdots K(s_2, \sigma_h) \\ \cdots\cdots\cdots\cdots\cdots & \cdots\cdots\cdots\cdots\cdots \\ K(s_m, t_1) \cdots K(s_m, t_m) & K(s_m, \sigma_1) \cdots K(s_m, \sigma_h) \\ K(\sigma_1, t_1) \cdots K(\sigma_1, t_m) & K(\sigma_1, \sigma_1) \cdots K(\sigma_1, \sigma_h) \\ \cdots\cdots\cdots\cdots\cdots & \cdots\cdots\cdots\cdots\cdots \\ K(\sigma_h, t_1) \cdots K(\sigma_h, t_m) & K(\sigma_h, \sigma_1) \cdots K(\sigma_h, \sigma_h) \end{vmatrix} d\sigma_1 \cdots d\sigma_h.$$

[1] See Fredholm, loc. cit.

We add the remark that the null solutions are obtained for the zeros $\lambda = \lambda_i$ of $D(\lambda)$, in the case of simple poles, by determining the residues of the resolvent $K(s, t; \lambda)$ at these points. The proof follows easily from our formulas.[1]

§8. *Another Derivation of the Theory*

The general theory of integral equations as formulated above is based on the fact that from the family of solutions of the approximating integral equations it is possible to select a sequence which converges uniformly to a solution of the given integral equation. Nevertheless, the concepts of measure of independence and of asymptotic dimension of a sequence of functions, introduced in Ch. II, §3 enable us to base the theory of integral equations on a somewhat different foundation. Since this procedure provides some relevant additional insight it will be presented here.

1. A Lemma. *Let $\psi_1(s)$, $\psi_2(s)$, $\cdots$, be a sequence of functions whose norms are all below a fixed bound M and for which the relation*

$$\psi_n(s) - \lambda \int K(s, t)\psi_n(t)\, dt \Rightarrow 0 \tag{81}$$

holds in the sense of uniform convergence. Then the functions $\psi_n(s)$ form a smooth sequence of functions with finite asymptotic dimension r.

To prove this we note that relation (81) remains valid if the functions $\psi_n(s)$ are replaced by any functions $\chi_n(s)$ which are of the form $\chi_n(s) = x_1\psi_{n_1} + x_2\psi_{n_2} + \cdots + x_p\psi_{n_p}$. The absolute values of the coefficients $x_1, x_2, \cdots, x_p$ are assumed to be bounded, and the functions

$$\psi_{n_1}, \psi_{n_2}, \cdots, \psi_{n_p}$$

are any p functions of the sequence ψ_n such that the indices n_i tend to infinity with n. Now if, among the functions $\psi_n(s)$, there are sets of r functions each with arbitrarily large indices n such that the measure of independence of each of these sets remains above a fixed bound α, if, in other words, the asymptotic dimension of the sequence is at least r, then we may orthogonalize the functions of these sets; the coefficients that arise in this orthogonalization process will all be less than the bound $1/\sqrt{\alpha}$ (Ch. II, §3, 1). We

[1] For further details concerning the formal apparatus of the Fredholm theory see, for example, G. Kowalewski, Einführung in die Determinantentheorie, Veit, Leipzig, 1909.

thus obtain sets of r mutually orthogonal functions $\omega_{n,i}(s)$ $(i = 1, 2, \cdots, r; n = 1, 2, \cdots)$ for which the limit equation

$$\lim_{n\to\infty}\left(\omega_{n,i}(s) - \lambda \int K(s, t)\omega_{n,i}(t)\, dt\right) = 0 \tag{82}$$

holds uniformly in s. The usual argument using Bessel's inequality[1] leads to the result

$$\iint [K(s, t)]^2\, ds\, dt \geq \sum_{i=1}^{r} \int \left[\int K(s, t)\omega_{n,i}(t)\, dt\right]^2 ds$$

for every n; and from this and from equation (82) it follows that

$$\iint [K(s, t)]^2\, ds\, dt \geq \frac{r}{\lambda^2}.$$

We have thus found a finite bound for the asymptotic dimension of the sequence; that the sequence is smooth follows immediately from (82). For, if ϵ_n denotes a number which tends to zero as n increases, the Schwarz inequality leads to

$$[\psi_n(s)]^2 \leq M\lambda^2 \int [K(s, t)]^2\, dt + \epsilon_n$$

which means that the functions $\psi_n(s)$ are uniformly bounded; similarly the relation

$$(x_1\psi_{n_1} + \cdots + x_p\psi_{n_p})^2 \leq \epsilon\lambda^2 \int [K(s, t)]^2\, dt + \epsilon_n$$

follows from $\int (x_1\psi_{n_1} + \cdots + x_p\psi_{n_p})^2\, ds < \epsilon$. Therefore the sequence is smooth.

2. Eigenfunctions of a Symmetric Kernel. We shall employ the lemma just proved to obtain the eigenfunctions of a symmetric kernel $K(s, t)$ which we approximate uniformly by the degenerate symmetric kernels $A_n(s, t)$. As before, let $\mu_1^{(n)}$, $\mu_2^{(n)}, \cdots$ and $\mu_{-1}^{(n)}$, $\mu_{-2}^{(n)}, \cdots$ denote the positive and negative eigenvalues, respectively, of $A_n(s, t)$, and let $\psi_1^{(n)}(s)$, $\psi_2^{(n)}(s), \cdots, \psi_{-1}^{(n)}(s)$, $\psi_{-2}^{(n)}(s), \cdots$ be the corresponding eigenfunctions. Multiple eigenvalues are supposed to be written down an appropriate number of times. Let

$$J_n(\varphi, \varphi) = \iint A_n(s, t)\varphi(s)\varphi(t)\, ds\, dt$$

[1] See §4, 2 of this chapter (p. 130).

and

$$J(\varphi, \varphi) = \iint K(s, t)\varphi(s)\varphi(t)\, ds\, dt$$

be the integral forms belonging to A_n and K, respectively, and let us assume, as we may, that $J(\varphi, \varphi)$ admits positive values. Then $\kappa_1^{(n)} = 1/\mu_1^{(n)}$ is the maximum of $J_n(\varphi, \varphi)$ for a normalized function; let $\kappa_1 = 1/\mu_1$ be the upper bound of $J(\varphi, \varphi)$ under the same normalization condition. Since, for sufficiently large n, the values of $J(\varphi, \varphi)$ and $J_n(\varphi, \varphi)$ differ by less than any arbitrarily small fixed number, we must have $\lim_{n\to\infty} \mu_1^{(n)} = \mu_1$. Therefore relation

$$\psi_1^{(n)}(s) - \mu_1 \int K(s, t)\psi_1^{(n)}(t)\, dt \Longrightarrow 0 \tag{83}$$

follows from $\psi_1^{(n)}(s) - \mu_1^{(n)} \int A_n(s, t)\psi_1^{(n)}(t)\, dt = 0$ since $A_n(s, t) \Longrightarrow K(s, t)$. Consequently, in virtue of our lemma, the functions $\psi_1^{(n)}$ form a smooth sequence of finite positive dimension r (the vanishing of r would contradict the normalization of the functions $\psi_1^{(n)}$). According to Ch. II, §3 they therefore determine a linear space of functions with the normalized orthogonal components $\psi_{1,1}(s)$, $\psi_{1,2}(s)$, $\cdots$, $\psi_{1,r}(s)$, which are necessarily solutions of the homogeneous integral equation

$$\psi_{1,i}(s) = \mu_1 \int K(s, t)\psi_{1,i}(t)\, dt \qquad (i = 1, 2, \cdots, r);$$

thus the functions $\psi_1^{(n)}$ are eigenfunctions of $K(s, t)$ belonging to the eigenvalue μ_1.

In exactly the same way we may obtain the remaining eigenvalues and eigenfunctions of $K(s, t)$. Thus, for example, $\kappa_h^{(n)} = 1/\mu_h^{(n)}$ is the minimum, obtained by a suitable choice of $v_1(s)$, $v_2(s)$, $\cdots$, $v_{h-1}(s)$, of the maximum of $J_n(\varphi, \varphi)$ subject to the conditions $(\varphi, \varphi) = 1$ and $(\varphi, v_i) = 0$ $(i = 1, 2, \cdots, h - 1)$.

If we now define $\kappa_h = 1/\mu_h$ as the analogous lower bound of the upper bound of $J(\varphi, \varphi)$, then we again have $\lim_{n\to\infty} \mu_h^{(n)} = \mu_h$ because of the proximity of the values of $J_n(\varphi, \varphi)$ to those of $J(\varphi, \varphi)$. This leads to the relation

$$\psi_h^{(n)}(s) - \mu_h \int K(s, t)\psi_h^{(n)}(t)\, dt \Rightarrow 0,$$

and the rest of the argument proceeds as above. To obtain the negative eigenvalues and the corresponding eigenfunctions, we must consider the corresponding minimum or maximum-minimum problems. If only a finite number of eigenvalues of one or the other sign occur we simply break off the procedure at the appropriate stage.

3. Unsymmetric Kernels. This method also provides simplification for the unsymmetric integral equation (1). A brief indication will suffice; we employ the same notation as before. In Case I, suppose the quantities ρ_n and c_n are such that for all n the norm c_n^2 remains below the bound M. Then the norm of the difference $\rho_n - \rho_m = \zeta_{nm}$ also remains bounded, namely less than $4M$. Furthermore

$$\lim_{n,m\to\infty} \left[\zeta_{nm}(s) - \lambda \int K(s, t)\zeta_{nm}(t)\, dt \right] = 0$$

uniformly in s. Therefore, by our lemma, every subsequence of the double sequence ζ_{nm} in which n and m both tend to infinity possesses a bounded asymptotic dimension r, and the bound of r depends only on $K(s, t)$ and λ. Thus our double sequence ζ_{nm} defines, through a limiting process, a linear space of functions (see Ch. II, §3), with a finite number r of orthorgonal components $\psi_1(s)$, $\psi_2(s)$, $\cdots$, $\psi_r(s)$, unless the asymptotic dimension of every subsequence is zero, i.e. unless $\zeta_{nm} \Longrightarrow 0$. In the latter case the $\rho_n(s)$ simply converge uniformly to a solution of the integral equation (1). In the case $r > 0$ the $\psi_i(s)$ are solutions of the homogeneous equation. We replace ρ_n by a function

$$\eta_n(s) = \rho_n(s) + x_1\psi_1(s) + \cdots + x_r\psi_r(s)$$

which is orthogonal to $\psi_1(s)$, $\psi_2(s)$, $\cdots$, $\psi_r(s)$. For these functions the relation

$$\left[\eta_n(s) - \lambda \int K(s, t)\eta_n(t)\, dt \right] - f(s) \Longrightarrow 0$$

certainly holds. As above, we may now apply the lemma to the differences $\eta_n - \eta_m = \zeta_{nm}$; we arrive easily at the conclusion that the dimension of every partial subsequence of this sequence must be zero, i.e. that the functions $\eta_n(s)$ converge uniformly to a solution of the integral equation orthogonal to the functions $\psi_i(s)$.

Similarly, in Case II we obtain, with the aid of the lemma, a linear

space of solutions of the homogeneous integral equation as the limit set of the sequence $\sigma_n(s) = \rho_n(s)/c_n$.

In this way the second approach provides more precise information concerning the nature of the convergence relations which obtain here. In particular, we see that a solution of the homogeneous or inhomogeneous equation may be approximated as closely as desired by considering approximating integral equations with the kernels $A_n(s, t)$.

4. Continuous Dependence of Eigenvalues and Eigenfunctions on the Kernel. In considering how the solutions of integral equations vary with the kernel, we restrict ourselves to the eigenvalue problem for a symmetric kernel $K(s, t)$. Suppose that the kernel $K(s, t)$ is the uniform limit of a sequence of other symmetric kernels $K_n(s, t)$ $(n = 1, 2, \cdots)$. If we consider functions φ for which $(\varphi, \varphi) \leq M$, then the values of the quadratic integral forms $J_n(\varphi, \varphi)$ and $J(\varphi, \varphi)$, with the kernels K_n and K, respectively, will differ arbitrarily little for sufficiently large n. The same is therefore true of the maxima and minima of these forms under the conditions $(\varphi, \varphi) = 1$, $(\varphi, v_i) = 0$, and also for the minima of the maxima or the maxima of the minima. In other words: *The h-th positive and h-th negative eigenvalues vary continuously with the kernel.* As far as the eigenfunctions are concerned, we cannot expect actual continuity in view of the arbitrariness of sign and of the occurrence of multiple eigenvalues. Instead, the following is true: *Let λ_h be an r-fold eigenvalue of the kernel $K(s, t)$, i.e. let*

$$\lambda_h = \lim \lambda_h^{(n)} = \lim \lambda_{h+1}^{(n)} = \cdots = \lim \lambda_{h+r-1}^{(n)},$$

but assume that this relation does not hold for $\lambda_{h-1}^{(n)}$ and $\lambda_{h+r}^{(n)}$. Then, for increasing n, the linear space formed from the eigenfunctions $\psi_h^{(n)}(s), \psi_{h+1}^{(n)}(s), \cdots, \psi_{h+r-1}^{(n)}(s)$ of the kernel $K_n(s, t)$ converges[1] uniformly to the linear space of the eigenfunctions of $K(s, t)$ for the eigenvalue λ_h.

This theorem is a complete expression of the continuity properties of the eigenfunctions. It may be proved almost immediately on the basis of our lemma by noting that, for the sequence of eigenfunctions $\psi_{h+k}^{(n)}(s)$ $(0 \leq k < r)$, the limit relation

$$\left[\psi_{h+k}^{(n)}(s) - \lambda_h \int K(s, t)\psi_{h+k}^{(n)}(t)\, dt\right] \Longrightarrow 0$$

holds, and that this sequence is certainly of asymptotic dimension r.

[1] For the notion of convergence of linear families see Ch. II, §3, 2.

§9. *Extensions of the Theory*

The developments of §§1 to 6 and of §8 can be significantly generalized in two directions.

In the first place, all the arguments remain valid if we consider integral equations for functions of several, say m, independent variables. Thus, suppose $f(s)$ and $\varphi(s)$ are continuous functions of the variables $s_1, s_2, \cdots, s_m$ in a fixed finite region G; suppose also that $K(s, t)$ is a continuous function of the variables $s_1, s_2, \cdots, s_m$ and $t_1, t_2, \cdots, t_m$, each set of variables ranging over the region G; let ds denote the volume element $ds_1 ds_2 \cdots ds_m$ of G and dt the corresponding volume element $dt_1 dt_2 \cdots dt_m$ with the understanding that all integrals are to be taken over the region G. Then the integral equation

$$f(s) = \varphi(s) - \lambda \int K(s, t)\varphi(t)\, dt$$

is an integral equation for a function $\varphi(s)$ of m variables, and its kernel $K(s, t)$ is a function of $2m$ variables; our whole theory remains valid word for word.

Secondly, the requirement that the kernel be continuous is far more stringent than necessary. It can be considerably relaxed without impairing the results. We shall not attempt to obtain the most general conditions under which the theory is valid; we shall merely extend the theory sufficiently far to treat the most important applications. Let us first consider kernels $K(s, t)$ which are functions of only two variables. Our previous arguments, apart from those that lead to Mercer's theorem (§5, 4), remain valid, with inessential modifications, for piecewise continuous kernels, since, as shown in Chapter I, any piecewise continuous function can be approximated in the mean as closely as desired by a continuous function. The arguments can also be justified, moreover, for kernels which are infinite at some points, provided that the integrals

$$\iint [K(s, t)]^2\, ds\, dt, \qquad \int [K(s, t)]^2\, ds, \qquad \int [K(s, t)]^2\, dt$$

exist and that the latter two integrals, regarded as functions of t and of s, respectively, remain below a fixed bound. The latter assumptions will be satisfied if, for example, the kernel becomes infinite for $s = t$ of order less that $\frac{1}{2}$, i.e. if $K(s, t)$ is of the form $K(s, t) =$

$H(s, t) \mid s - t \mid^{-\alpha}$ with $0 \leq \alpha < \frac{1}{2}$ and $H(s, t)$ continuous everywhere. To show that for such a kernel the theorems obtained are valid, it is sufficient to approximate the kernel by continuous degenerate kernels $A_n(s, t)$ in such a way that the following conditions are satisfied: $\int [K(s, t) - A_n(s, t)]^2 \, dt$ becomes arbitrarily small uniformly in s as $n \to \infty$ and $\int [A_n(s + \eta, t) - A_n(s, t)]^2 \, dt$ becomes arbitrarily small uniformly in s and in n if $\mid \eta \mid$ is taken sufficiently small. All our arguments can then be carried out. Likewise, in the case of two independent variables, all our theorems remain valid if, for $s_1 = t_1$, $s_2 = t_2$, the kernel becomes infinite of less than the first order, since in this case the convergence of the integral $\iint [K(s_1, s_2, t_1, t_2)]^2 \, ds_1 ds_2$ is not affected. Analogously, we may admit singularities of order less than $\frac{3}{2}$ in the case of three variables; in general for n variables singularities of order less than $n/2$ are permissible in K.

We finally note without proof that it is not difficult to extend our theory to integral equations for which only the above hypothesis concerning the integrals of $K^2(s, t)$ are satisfied. No additional requirements such as continuity of the kernel are needed.

§10. *Supplement and Problems for Chapter III*

1. Problems. Show that:

(a) The kernel

$$\sum_{n=1}^{\infty} \frac{\sin ns \sin nt}{n} = \tfrac{1}{2} \log \left| \sin \frac{s + t}{2} \Big/ \sin \frac{s - t}{2} \right| \qquad (0 \leq s, t \leq \pi)$$

has the eigenvalues $\lambda_n = 2n/\pi$ and the eigenfunctions $\sin nt$.

(b) The symmetric kernel

$$\frac{1}{2\pi} \frac{1 - h^2}{1 - 2h \cos (s - t) + h^2} \qquad (0 \leq s, t \leq 2\pi)$$

with $\mid h \mid < 1$ has the eigenfunctions 1, $\sin ns$, $\cos ns$, with the eigenvalues 1, $1/h^n$, $1/h^n$.

(c) For the symmetric kernel

$$K(s, t) = \frac{1}{\sqrt{\pi}} e^{(s^2+t^2)/2} \int_{-\infty}^{s} e^{-\tau^2} \, d\tau \int_{t}^{\infty} e^{-\tau^2} \, d\tau \qquad (s \leq t)$$

the Hermite orthogonal functions $e^{s^2/2}(d^n e^{-s^2}/ds^n)$ are eigenfunctions with the eigenvalues $\lambda_n = 2n + 2$.

(d) The eigenfunctions of the symmetric kernel

$$K(s, t) = e^{(s+t)/2} \int_t^\infty \frac{e^{-\tau}}{\tau}\, d\tau \qquad (0 \leq s \leq t)$$

are the Laguerre orthogonal functions

$$e^{-s/2} \frac{\partial^n}{\partial h^n} \frac{e^{-sh/(1-h)}}{1-h}\bigg|_{h=0};$$

the associated eigenvalues are $\lambda_n = n + 1$.[1]

2. Singular Integral Equations. The general theory may no longer be valid if the kernel possesses singularities of too high an order, or if, in the case of an infinite fundamental domain, it fails to go to zero fast enough at infinity (the kernels considered in the previous subsection converge strongly enough).

We shall now give some examples of integral equations with eigenvalues of infinite multiplicity.

The integral formula

$$\int_0^\infty \sin st \left(\sqrt{\frac{\pi}{2}}\, e^{-at} + \frac{t}{a^2 + t^2}\right) dt = \sqrt{\frac{\pi}{2}}\, e^{-as} + \frac{s}{a^2 + s^2}$$

holds identically in a. Therefore, in the fundamental domain $0 \leq s, t < \infty$, the kernel $\sin st$ has infinitely many eigenfunctions for the eigenvalue $\lambda = 1$.

Hermite's orthogonal functions (see subsection 1(c)) are eigenfunctions of the kernel e^{ist} with the eigenvalues $i^{-n}/\sqrt{2\pi}$. Thus each of the four numbers $\pm 1/\sqrt{2\pi}$, $\pm i/\sqrt{2\pi}$ is an eigenvalue of this kernel of infinite multiplicity.

An example of an integral equation[2] with infinitely many eigenvalues in a finite interval is the equation

$$\varphi(s) = \lambda \int_{-\infty}^{\infty} e^{-|s-t|}\, \varphi(t)\, dt;$$

[1] See R. Neumann, Die Entwicklung willkürlicher Funktionen nach den Hermiteschen und Laguerreschen Orthogonalfunktionen usw., Dissertation, Breslau, 1912.

[2] Related integral equations have been treated by E. Hopf: Über lineare Integralgleichungen mit positivem Kern, Sitzungsber. Akad. Berlin (math.-phys. Kl.), 1928, pp. 233–275. See also the articles by U. Wegner, G. H. Hardy, and E. C. Titchmarsh cited there.

its solutions are $e^{\alpha i s}$ with the eigenvalues $\lambda = \frac{1}{2}(1 + \alpha^2)$. Thus every $\lambda > \frac{1}{2}$ is an eigenvalue.

3. E. Schmidt's Derivation of the Fredholm Theorems.[1] Let us take $\lambda = 1$ and let us write the kernel $K(s, t)$ in the form

$$K(s, t) = \sum_{\nu=1}^{n} \alpha_\nu(s)\beta_\nu(t) + k(s, t),$$

with $\iint [k(s, t)]^2 ds\, dt < 1$. According to §6, the Neumann series for $k(s, t)$ converges for $\lambda = 1$ and therefore represents the resolvent $\kappa(s, t)$ of the kernel $k(s, t)$. Writing the integral equation (1) in the form

$$f_1(s) = \varphi(s) - \int k(s, t)\varphi(t)\, dt$$

with

$$f_1(s) = f(s) + \sum_{\nu=1}^{n} x_\nu\alpha_\nu(s) \qquad (x_\nu = (\varphi, \beta_\nu))$$

we thus have the formula

$$\varphi(s) = f(s) + \sum_{\nu=1}^{n} x_\nu\alpha_\nu(s) + \int \kappa(s, t)\left[f(t) + \sum_{\nu=1}^{n} x_\nu\alpha_\nu(t)\right] dt$$

or

$$f_2(s) = f(s) + \int \kappa(s, t)f(t)\, dt$$

$$= \varphi(s) - \int \left[\sum_{\nu=1}^{n} (\alpha_\nu(s)\beta_\nu(t) + \gamma_\nu(s)\beta_\nu(t))\right] \varphi(t)\, dt,$$

with

$$\gamma_\nu(s) = \int \kappa(s, \tau)\alpha_\nu(\tau)\, d\tau.$$

In this way, the given integral equation is reduced in a single step to one with a degenerate kernel.

[1] E. Schmidt, Zur Theorie der linearen und nicht linearen Integralgleichungen. Part II: Auflösung der allgemeinen linearen Integralgleichung, Math. Ann., Vol. 64, 1907, pp. 161–174.

4. Enskog's Method for Solving Symmetric Integral Equations.[1] We consider a positive definite kernel $K(s, t)$, the first eigenvalue of which is greater than 1, i.e. a kernel for which the inequality

$$\int [\varphi(s)]^2 \, ds - \iint K(s, t)\varphi(s)\varphi(t) \, ds \, dt > 0$$

is valid for all φ. The integral equation (1) (assuming $\lambda = 1$) may be written in the form $f(s) = J(\varphi)$, where

$$J(\varphi) = \varphi(s) - \int K(s, t)\varphi(t) \, dt.$$

From any complete system of functions $\varphi_1, \varphi_2, \cdots$, we now construct, by a process analogous to the orthogonalization process described in Chapter II, a system of functions $v_1(s), v_2(s), \cdots$ with the property that $\int v_i \, J(v_k) \, ds = \delta_{ik}$.[2] Such a system is known as a "complete system polar with respect to the kernel $K(s, t)$." If we set $a_\nu = \int \varphi \, J(v_\nu) \, ds = \int v_\nu \, f \, ds$, we immediately obtain $\varphi(s) = \sum_{\nu=1}^{\infty} a_\nu v_\nu(s)$, provided that this series converges uniformly. Incidentally, the functions v_ν satisfy the "completeness relation"

$$\int \varphi(s) J[\varphi(s)] \, ds = \sum_{\nu=1}^{\infty} a_\nu^2,$$

no matter how the piecewise continuous function $\varphi(s)$ is chosen.

5. Kellogg's Method for Determination of Eigenfunctions.[3] Starting with an arbitrary normalized function $\varphi_0(s)$, we determine the functions $\varphi_\nu(s)$ and the numbers λ_ν from the relations $\varphi_{\nu+1}(s) = \lambda_{\nu+1} \int K(s, t)\varphi_\nu(t) \, dt$, $N\varphi_\nu = 1$. The passage to the limit can be carried out and leads to an eigenvalue and associated eigenfunction of the kernel or its iterated kernel.

The reader may establish the connection between this approach and the notion of asymptotic dimension and carry out the derivation from this point of view.

[1] D. Enskog, Kinetische Theorie der Vorgänge in mässig verdünnten Gasen, Dissertation, Uppsala, 1917.

[2] δ_{ik} is the "Kronecker delta" defined by $\delta_{ii} = 1$, $\delta_{ik} = 0$ for $i \neq k$.

[3] O. D. Kellogg, On the existence and closure of sets of characteristic functions, Math. Ann., Vol. 86, 1922, p. 14–17.

6. Symbolic Functions of a Kernel and Their Eigenvalues. For the integral operator associated with a given kernel, spectral relations analogous to those developed for matrices in Chapter I exist. Let us consider, in particular, an integral rational function $f(u) = \sum_{\nu=1}^{n} a_\nu u^\nu$ which vanishes for $u = 0$; we replace the powers of u by the corresponding iterated kernels of the symmetric kernel $K(s, t)$ and thus obtain the kernel

$$H(s, t) = f[K] = \sum_{\nu=1}^{n} a_\nu K^{(\nu)}(s, t).$$

The following theorem now holds: *The eigenfunctions φ_i of H are identical with the eigenfunctions of K, and the corresponding reciprocal eigenvalues η_i of H are connected with the reciprocal eigenvalues κ_i of K by the equation*

$$\eta_i = f(\kappa_i).$$

We may, in fact, verify immediately that each eigenfunction φ_i of K belonging to the eigenvalue $\lambda_i = 1/\kappa_i$ is also an eigenfunction of H with the eigenvalue $1/f(\kappa_i)$. Since the relation

$$\iint [H(s, t)]^2 \, ds \, dt = \sum_{i=1}^{\infty} [f(\kappa_i)]^2$$

holds we see that H has no other eigenvalues and eigenfunctions.

7. Example of an Unsymmetric Kernel without Null Solutions. The kernel

$$K(s, t) = \sum_{\nu=1}^{\infty} \frac{\sin \nu s \sin (\nu + 1)t}{\nu^2}$$

has, in the region $0 \leq s, \quad t \leq 2\pi$, no null solutions. For, the iterated kernels are

$$K^{(n)}(s, t) = \pi^{n-1} \sum_{\nu} \frac{\sin \nu s \sin (\nu + n)t}{\nu^2(\nu + 1)^2 \cdots (\nu + n - 1)^2},$$

and therefore the Neumann series converges for all values of λ. The same result may be obtained by showing that the function $D(\lambda)$ associated with K is a constant for this kernel.[1]

[1] Similar kernels may be found in Goursat, Cours d'analyse (see bibliography).

8. Volterra Integral Equation.[1] If $K(s, t) = 0$ for $s < t$, the integral equation may be written in the form

$$f(s) = \varphi(s) - \lambda \int_a^s K(s, t)\phi(t)\, dt.$$

Integral equations of this type have been treated in particular by Volterra. The reader may show that the resolvent for this equation is an integral transcendental function of λ, and that therefore Volterra's integral equation possesses one and only one solution for every λ, consequently, no null solution for any λ and its Neuman series converges.

9. Abel's Integral Equation.[2] Abel studied a particular integral equation of the Volterra type, which is important in many applications, in order to solve the following problem: Let a material point of mass m move under the influence of gravity on a smooth curve lying in a vertical plane. Let the time t which is required for the point to move along the curve from the height x to the lowest point of the curve be a given function f of x. What is the equation of the curve? This problem leads to the integral equation

$$f(x) = \int_0^x \frac{\varphi(t)\, dt}{\sqrt{2g(x - t)}}.$$

If we assume that $f(x)$ is a continuously differentiable function vanishing at $x = 0$, the solution of the equation turns out to be

$$\varphi(x) = \frac{\sqrt{2g}}{\pi} \int_0^x \frac{f'(t)\, dt}{\sqrt{x - t}},$$

where g is the acceleration of gravity; the equation of the curve is

$$y = \int_0^x \sqrt{|\, \varphi^2(t) - 1\,|}\, dt.$$

More generally, we may consider the equation

$$f(x) = \int_a^x \frac{\varphi(s)\, ds}{(s - x)^\alpha} \qquad (0 < \alpha < 1),$$

[1] V. Volterra, Leçons sur les équations intégrales et les équations intégro-différentielles, Chapter II, Gautier-Villars, Paris, 1913.

[2] Abel, Solution de quelques problèmes à l'aide d'intégrales définies, Works, Christiania, 1881, I, pp. 11–27; Bôcher, Integral Equations, p. 8, Cambridge University Press, Cambridge, 1909.

the solution of which, for continuously differentiable $f(x)$, is

$$\varphi(x) = \frac{\sin \alpha\pi}{\pi} \frac{f(a)}{(x-a)^{1-\alpha}} + \frac{\sin \alpha\pi}{\pi} \int_a^x \frac{f'(s)\, ds}{(s-x)^{1-\alpha}}.$$

10. Adjoint Orthogonal Systems belonging to an Unsymmetric Kernel.[1] From an unsymmetric kernel $K(s, t)$ we may construct the two symmetric kernels $K'(s, t) = \int K(s, \sigma)K(t, \sigma)\, d\sigma$ and $K''(s, t) = \int K(\sigma, s)K(\sigma, t)\, d\sigma$. Then there exists a sequence of pairs of functions and associated values λ such that

$$\varphi_\nu(s) = \lambda_\nu \int K(s, t)\psi_\nu(t)\, dt, \qquad \psi_\nu(s) = \lambda_\nu \int K(t, s)\varphi_\nu(t)\, dt,$$

$$\varphi_\nu(s) = \lambda_\nu^2 \int K'(s, t)\varphi_\nu(t)\, dt, \qquad \psi_\nu(s) = \lambda_\nu^2 \int K''(s, t)\psi_\nu(t)\, dt.$$

Every function which can be written in the form $\int K(s, t)h(t)\, dt$ admits of an absolutely and uniformly convergent expansion in a series of functions of the orthogonal system φ_ν; similarly, any function of the form $\int K(t, s)h(t)\, dt$ can be expanded in a series of the functions ψ_ν. The relation $K(s, t) = \sum_{\nu=1}^{\infty} \varphi_\nu(s)\psi_\nu(t)/\lambda_\nu$ holds, provided the series on the right converges uniformly in each variable. The kernel K is determined uniquely by the values λ_ν and by the two independent orthogonal systems.

11. Integral Equations of the First Kind. An integral equation of the first kind is an equation of the form

$$f(s) = \int K(s, t)\varphi(t)\, dt. \tag{84}$$

We have seen, for example, that a function can be expanded in a series of the eigenfunctions of a kernel if a solution of the integral equation of the first kind (53) exists. Other examples are given by the Fourier integral and the Mellin integral transformation (Ch. II, §10, 8). The specific difficulty of the theory of integral equations of the first kind is due to the following fact: for a continuous kernel $K(s, t)$, the equation (84) transforms the manifold of all piecewise

[1] E. Schmidt, Zur Theorie der linearen und nichtlinearen Integralgleichungen. I. Entwicklung willkürlicher Funktionen nach Systemen vorgeschriebener. Math. Ann. Vol. 63, 1907, pp. 433–476.

continuous functions $\varphi(s)$ into a more restrictive manifold, since all functions $f(s)$ obtained this way are certainly continuous. If $K(s, t)$ is differentiable every piecewise continuous function, and in fact every integrable function, is transformed into a differentiable function. Thus, in general, the integral equation with continuous $f(s)$ cannot be solved by a continuous function φ. For more general classes of functions $f(s)$ we may expect (84) to be solvable only if $K(s, t)$ deviates from regular behavior in some way. The reader may consider the preceding and subsequent examples of such integral equations from this point of view; an infinite fundamental domain has the same effect as a singularity of the kernel.

For a symmetric kernel, one may attempt a solution of the form $\varphi(s) = \sum_{\nu=1}^{\infty} \lambda_\nu x_\nu \varphi_\nu(s)$, where $x_\nu = (f, \varphi_\nu)$ are the Fourier expansion coefficients of f with respect to the system $\varphi_1, \varphi_2, \cdots$ of eigenfunctions of the kernel. If this series converges uniformly—a condition that restricts $f(s)$ in view of the increase in λ_ν with increasing ν—then it actually provides a solution of (84).

The general case is covered by a theorem of Picard[1] which gives necessary and sufficient conditions for the existence of a square integrable solution $\varphi(s)$ of an equation of the first kind $f(s) = \int K(s, t)\varphi(t)\, dt$ with an arbitrary (possibly unsymmetric) kernel K: If φ_i, ψ_i are the pairs of adjoint functions belonging to $K(s, t)$, as defined in subsection 10, and λ_i the corresponding eigenvalues, then the above integral equation can be solved if and only if the series

$$\sum_{i=1}^{\infty} \lambda_i^2 \left(\int f(s)\varphi_i(s)\, ds \right)^2$$

converges.

12. Method of Infinitely Many Variables. If $\omega_1(s)$, $\omega_2(s)$, $\cdots$ is a complete orthogonal system in the fundamental domain, and if we define $x_i = (\varphi, \omega_i)$, $f_i = (f, \omega_i)$, and $k_{pq} = \iint K(s, t)\omega_p(s)\omega_q(t)\, ds\, dt$, then the integral equation (1) leads immediately to the system

$$f_i = x_i - \lambda \sum_{j=1}^{\infty} k_{ij} x_j \qquad (i = 1, 2, \cdots)$$

of infinitely many linear algebraic equations for the infinitely

[1] E. Picard, Sur un théorème général relatif aux équations intégrales de première espèce et sur quelques problèmes de physique mathématique. Rend. Circ. math. Palermo, V. 29, 1910, pp. 79–97.

many unknowns x_1, x_2, $\cdots$. The sums $\sum_{i=1}^{\infty} x_i^2$, $\sum_{i=1}^{\infty} f_i^2$, and $\sum_{i,j=1}^{\infty} k_{ij}^2$ converge, as may be deduced from Bessel's inequality. The theory of the solutions of this system of equations then leads to the theorems about the integral equation (1).

13. Minimum Properties of Eigenfunctions. The eigenfunctions φ_1, φ_2, $\cdots$ of a symmetric kernel, or the two orthogonal systems φ_i and ψ_i for an unsymmetric kernel, and the corresponding eigenvalues λ_i, may be obtained from the following minimum problem: The kernel $K(s, t)$ is to be approximated by a degenerate kernel $A_n(s, t) = \sum_{i=1}^{n} \Phi_i(s)\Psi_i(t)/\Lambda_i$ in such a way that $\iint (K - A_n)^2 \, ds \, dt$ becomes as small as possible. The reader may prove that the solution is given by $\Phi_i = \varphi_i$, $\Psi_i = \psi_i$, $\Lambda_i = \lambda_i$.

14. Polar Integral Equations. For kernels of the form $K(s, t) = A(s)S(s, t)$, where $S(s, t)$ is symmetric and $A(s)$ is continuous except for a finite number of jumps, it is possible to derive results similar to those obtained for symmetric kernels. The case where $S(s, t)$ is definite, i.e. where it has only positive (or only negative) eigenvalues, has been studied extensively. In this case, which has been treated by Hilbert[1] and Garbe,[2] the integral equation is said to be *polar* or *of the third kind.* As was the case for symmetric kernels, the resolvent has only simple and real poles. For the corresponding residues, which yield the "polar eigenfunctions," an expansion development theorem holds which is analogous to that formulated by Hilbert for symmetric kernels. In particular, if the iterated kernel $K^{(2)}(s, t)$ does not vanish identically at least one eigenvalue exists. Incidentally, the theorem that the resolvent has only real simple poles remains valid if $S(s, t)$ is simply assumed to be *positive*; the additional theorem that there is at least one eigenvalue holds if $S(s, t)$ is positive and $K^{(2)}(s, t)$ does not vanish identically.[3]

15. Symmetrizable Kernels.[4] Kernels for which the resolvent has only real simple poles may be characterized in the following

[1] D. Hilbert, Integralgleichungen, Chapter 15. Here a somewhat different form is taken for the polar integral equation.

[2] E. Garbe, Zur Theorie der Integralgleichung dritter Art. Math. Ann., Vol. 76, 1915, pp. 527–547.

[3] J. Marty, Sur une équation intégrale. C. R. Acad. sc. Paris, Vol. 150, 1910, pp. 515–518. Développements suivant certaines solutions singulières, ibid., pp. 603–606. Existence de solutions singulières pour certaines équations de Fredholm, ibid., pp. 1031–1033.

[4] J. Marty, Valeurs singulières d'une équation de Fredholm, C. R. Acad. sc. Paris, Vol. 150, 1910, pp. 1499–1502.

simple manner: In order that a kernel have this property it is necessary that there exist a kernel $S(s, t)$ such that the kernels $\int S(s, \tau)K(\tau, t)\,d\tau$ and $\int K(s, \tau)S(\tau, t)\,d\tau$ are symmetric. Such kernels $K(s, t)$ are said to be *symmetrizable*. Conversely, if for a suitable positive definite symmetric kernel $S(s, t)$ at least one of the above integrals represents a symmetric kernel, then all the poles of the resolvent of $K(s, t)$ are real and simple.

16. Determination of the Resolvent Kernel by Functional Equations. Prove that the resolvent of $K(s, t)$ is uniquely determined by the equations (63).

17. Continuity of Definite Kernels. Prove that, if a definite symmetric kernel $K(s, t)$ is piecewise continuous for $0 \leq s, t \leq 1$ and if it is continuous at all points $s = t$ and has continuous eigenfunctions, then it is continuous everywhere in the region $0 \leq s, t \leq 1$.

18. Hammerstein's Theorem. If the kernel $K(s, t)$ is continuous in the fundamental domain $0 \leq s, t \leq 1$ and has uniformly bounded first derivatives in this domain, then the bilinear formula holds for the kernel itself and not only for the iterated kernel $K^{(2)}(s, t)$. The requirement of bounded differentiability can actually be replaced by considerably less restrictive conditions.[1]

References

We refer particularly to the article by E. Hellinger and O. Toeplitz in the Enzyklopädie der mathematischen Wissenschaften, Vol. 2. This article contains a unified exposition of the theory of integral equations and deals in detail with the connection between this theory and other fields of analysis. We may further refer to the review article by H. Hahn, Bericht über die Theorie der linearen Integralgleichungen, Jahresber. d. deutschen Mathematiker-Vereinigung, Vol. 20, 1911, pp. 69–117.

Textbooks

Bôcher, M., An introduction to the study of integral equations. Cambridge tracts, Vol. 10, Cambridge University Press, Cambridge 1909.

Goursat, E., Cours d'analyse mathématique, Vol. 3, 3rd ed., Gauthier-Villars, Paris 1923, pp. 323–544.

Kneser, A.: Die Integralgleichungen und ihre Anwendungen in der mathematischen Physik. 2nd ed., Vieweg, Braunschweig 1922.

[1] A. Hammerstein, Über die Entwicklung des Kernes linearer Integralgleichungen und Eigenfunktionen. Sitzungsber. Akad. Berlin (phys.-math. Kl.), 1923, pp. 181–184.

Kowalewski, G., Einführung in die Determinantentheorie, einschliesslich der unendlichen und der Fredholmschen Determinanten. De Gruyter, Leipzig, 1909.

Lalesco, T., Introduction à la théorie des équations intégrales. Hermann, Paris, 1912. (Containing a detailed bibliography up to 1912.)

Vivanti, G., Elementi della teoria delle equazioni integrali lineare. Mailand, 1916. (German edition F. Schwank, Hannover 1929.)

Volterra, V., Leçons sur les équations intégrales et les équations intégro-différentielles. Gauthier-Villars, Paris, 1913.

Monographs and Articles

Carleman, T., Sur les équations intégrales singulières à noyau réel et symmétrique. Uppsala Univ. Årsskrift 1923.

Courant, R., Zur Theorie der linearen Integralgleichungen. Math. Ann., Vol. 89, 1923, pp. 161–178.

Fredholm, I., Sur une classe d'équations fonctionnelles. Acta math., Vol. 27, 1903, pp. 365–390.

Goursat, E., Recherches sur les équations intégrales linéaires. Ann. Fac. sc. Toulouse, Series 2, Vol. 10, 1908, pp. 5–98.

Hilbert, D., Grundzüge einer allgemeinen Theorie der linearen Integralgleichungen. Leipzig and Berlin, 1912. (Reprint of six communications from the Nachrichten der K. Gesellschaft der Wissenschaften zu Göttingen, 1904–1910.)

Holmgren, E., Das Dirichletsche Prinzip und die Theorie der linearen Integralgleichungen. Math. Ann., Vol. 69, 1910, pp. 498–513.

Landsberg, G., Theorie der Elementarteiler linearer Integralgleichungen. Math. Ann., Vol. 69, 1910, p. 227–265.

Schmidt, E., Zur Theorie der linearen und nichtlinearen Integralgleichungen. Math. Ann., Vol. 63, 1907, pp. 433–476. Ibid., Vol. 64, 1907, pp. 161–174.

Schur, I., Über die charakteristischen Wurzeln einer linearen Substitution mit einer Anwendung auf die Theorie der Integralgleichungen. Math. Ann., Vol. 66, 1909, pp. 488–510.

Weyl, H., Singuläre Integralgleichungen mit besonderer Berücksichtigung des Fourierschen Integraltheorems. Dissertation, Göttingen, 1908.

CHAPTER IV

The Calculus of Variations

Many problems of mathematical physics are connected with the calculus of variations, one of the very central fields of analysis. In this chapter we shall develop certain fundamental concepts of the variational calculus in order to obtain the differential equations of mathematical physics and methods for their solution.

§1. *Problems of the Calculus of Variations*

1. Maxima and Minima of Functions. The calculus of variations originates from a generalization of the elementary theory of maxima and minima. For a better understanding we shall first examine the well-known elementary theory. It is concerned with the problem of finding, for a given continuous function $f(x, y, \cdots)$ in a given closed region G, a point $x_0, y_0 \cdots$ in G at which the function $f(x, y, \cdots)$ has a maximum or minimum (extremum) with respect to all points of G in a vicinity of $x_0, y_0, \cdots$. This problem always has a solution, according to the theorem of Weierstrass: *Every function which is continuous in a closed domain G of the variables possesses a largest and a smallest value in the interior or on the boundary of the domain.* If the function $f(x, y, \cdots)$ is differentiable in G and if an extreme value is attained at an interior point $x_0, y_0, \cdots$, then the derivatives of $f(x, y, \cdots)$ with respect to each variable vanish at $x_0, y_0, \cdots$; in other words, the gradient of f is equal to zero. But this necessary condition is in no way sufficient, as can be seen from the existence of points of inflection or saddle points (examples are $f(x) = x^3$ at $x_0 = 0$; $f(x, y) = xy$ at $x_0 = 0, y_0 = 0$). In general, points at which all the derivatives of the function vanish, so that $df = 0$, are known as *stationary points.* Stationary points which furnish a maximum or minimum relative to a suitable vicinity are called "extrema."

If the variables are not independent but are subject to the restrictions $g_1(x, y, \cdots) = 0$, $g_2(x, y, \cdots) = 0, \cdots, g_h(x, y, \cdots) = 0$,

we obtain necessary conditions for an extremum or stationary point by means of *Lagrange's method of multipliers*. This method consists in the following procedure: In order to find a point in the interior of the domain of the independent variables at which $f(x, y, \cdots)$ has an extremum or is merely stationary, we introduce $h + 1$ new parameters, the "multipliers," $\lambda_0, \lambda_1, \cdots, \lambda_h$ and construct the function $F = \lambda_0 f + \lambda_1 g_1 + \lambda_2 g_2 + \cdots + \lambda_h g_h$. We now determine the quantities $x_0, y_0, \cdots$ and the ratios of $\lambda_0, \lambda_1, \cdots, \lambda_h$ from equations

$$\begin{aligned} &\frac{\partial F}{\partial x} = 0, \qquad \frac{\partial F}{\partial y} = 0, \qquad \cdots \\ &\frac{\partial F}{\partial \lambda_1} = g_1 = 0, \qquad \cdots, \qquad \frac{\partial F}{\partial \lambda_h} = g_h = 0, \end{aligned} \tag{1}$$

the number of which is equal to the number of unknowns. These equations represent the desired conditions for the stationary character of $f(x, y, \cdots)$ or the extremum of f under the given restrictions.

If $\lambda_0 \neq 0$ we may (and shall) put $\lambda_0 = 1$ because F is homogeneous in the quantities λ_i. The Lagrange method is simply a device which, preserving symmetry, avoids explicit elimination of h of the variables from the function $f(x, y, \cdots)$ by means of the subsidiary restrictions.

We now consider a few instructive though elementary examples.

(a) *Of all triangles with given base line and given perimeter the isosceles triangle has the greatest area. Of all triangles with given base line and area the isosceles triangle has the least perimeter.* These statements may easily be verified without calculation by considering the ellipses for which the given base line is the line connecting the two foci. Even this simple example exhibits a characteristic *reciprocity*, which we shall encounter again later (§12, 2).

(b) *Refraction and Reflection of Light.* According to Fermat's *principle of least time* a ray of light requires less time along its actual path between two points than it would require along any other conceivable ("virtual") path satisfying the given conditions. From this it follows immediately that light travels in straight lines in any homogeneous medium. If the light ray is required to meet a given curve (mirror) without crossing it, then a simple analysis of the derivative of the light time shows that the path must consist of two straight line-segments making equal angles with the tangent to the given curve at the point where they meet it (*law of reflection*). If, on the other hand, the given curve forms the boundary between two

regions in which the velocity of light has the values c_1 and c_2, respectively, and if the ray is to pass from one region to the other, then it must consist of two straight line-segments which satisfy the well-known *law of refraction* (*Snell's law*), $\sin\alpha_1/\sin\alpha_2 = c_1/c_2$, where α_1 and α_2 are the angles between the normal to the boundary curve and the two line-segments at the point of intersection.

(c) *A Problem of Steiner.* Given three points A_1, A_2, A_3 which form an acute triangle, a fourth point P is to be found such that the sum of the distances $PA_1 + PA_2 + PA_3$ is as small as possible. We consider a circle through P with center at A_3; then P must be placed in this circle in such a way that $PA_1 + PA_2$ is as small as possible. According to the law of reflection this means that the lines PA_1 and PA_2 make equal angles with the radius PA_3. The same must be true if the indices 1, 2, 3 are interchanged; all three angles A_1PA_2, A_2PA_3, A_3PA_1 must, therefore, be equal and equal to $2\pi/3$. Thus the problem is solved.

(d) *The Isoperimetric Problem for Polygons.* Among all polygons which are not self-intersecting and have an even number $2n$ of sides and a given perimeter $2l$ find the one with the greatest area. The desired polygon $\Pi(A_1, A_2, \cdots, A_{2n})$ is the regular $2n$-gon. To prove this we first convince ourselves that Π is convex. If Π were not convex, a straight line lying entirely outside Π could be drawn through two non-adjacent vertices, say A_k and A_l; we could then reflect the polygonal sequence $A_kA_{k+1}\cdots A_{l-1}A_l$ in this straight line and obtain a polygon with the same perimeter and with an area greater than that of the original polygon by twice the area of the triangle $A_1A_2A_3$. We may thus confine ourselves to convex polygons. Next we show that all the sides of the polygon are equal in length. For, if the lengths of two successive sides A_1A_2 and A_2A_3 were different, we could, according to the result of (a), replace the vertex A_2 by a vertex A_2' such that $A_1A_2' + A_2'A_3 = A_1A_2 + A_2A_3$ while the area of the triangle $A_1A_2'A_3$ is greater than that of $A_1A_2A_3$. Thus the area of the polygon with the vertex A_2' is greater than that of the polygon with A_2, in contradiction to the hypothesis that Π is the maximal polygon. Finally, in order to show that the polygon Π can be inscribed in a circle, we divide Π into two polygons by means of the diagonal connecting opposite vertices A_1 and A_{n+1}. The perimeters of these polygons Π_1 and Π_2 are evidently equal; their areas are also equal, for if that of Π_1 were greater than that of Π_2 we could replace Π_2 by the reflection of Π_1 in the diagonal, obtaining a new polygon Π^*

whose perimeter is still $2l$ but whose area is twice that of Π_1 and therefore greater than that of Π. We now show that for every vertex A_k the angle $A_1A_kA_{n+1}$ must be a right angle. For, if this angle were not right for the vertex A_k of Π_1, we could decompose Π_1 into the triangle $A_1A_kA_{n+1}$ and two polygons H_1 and H_2 which join the triangle at the sides A_1A_k and A_kA_{n+1}. Consider a right triangle $A_1'A_kA_{n+1}'$ whose sides $A_1'A_k$ and A_kA_{n+1}' are equal to A_1A_k and A_kA_{n+1}, respectively. The area of this triangle is greater than that of $A_1A_kA_{n+1}$. If we affix the polygons H_1 and H_2 to the new triangle at its sides and reflect the resulting polygon in the line $A_1'A_{n+1}'$ we obtain a polygon of perimeter $2l$ and area greater than that of Π, contrary to hypothesis. Therefore $A_1A_kA_{n+1}$ is a right triangle, and consequently all vertices lie on a circle. This completes the proof of the extremum property of the regular polygon.[1]

The preceding treatment, based on a classical idea of Steiner, shows that in individual cases a direct geometrical approach may lead more readily to the result than general analytic procedures.

(e) *Other Examples, Maximum of a Minimum.* We have already encountered other typical examples of stationary values which are neither maxima nor minima but maxima of a minimum or minima of a maximum. For instance, we defined the eigenvalues of quadratic forms as maxima of minima and the Tchebycheff polynomials by a similar property.

2. Functionals. The calculus of variations likewise originates from the quest for extrema or stationary values. Its object is, however, to find extrema of *functionals* rather than extrema of functions of a finite number of independent variables. By a "functional" we mean a quantity or function which depends on the entire course of one or more functions rather than on a number of discrete variables. In other words, the domain of a functional is a set or "*space*" *of* admissible *functions* rather than a region of a coordinate space. A simple example is the length L of a curve $y = y(x)$ between the values $x = x_0$ and $x = x_1$. This length is given by the integral

$$L = \int_{x_0}^{x_1} \sqrt{1 + y'^2}\, dx;$$

[1] Note that the *existence* of the extremum follows from Weierstrass' theorem. For, if one vertex is at the origin the others are confined to a bounded region, since the total perimeter is given and the area of the polygon is a continuous function of the coordinates of the vertices.

thus the value of L depends on the course of the "*argument function*" $y(x)$, which may be taken to be an arbitrary continuous function with a piecewise continuous derivative. Such functionals occur throughout analysis and its applications, and many important problems of analysis are concerned with functional dependence of this kind.

Another example is the area of a surface $z = z(x, y)$ lying above the region G in the xy-plane. This is given by the integral

$$\iint_G \sqrt{1 + z_x^2 + z_y^2}\, dx\, dy^1$$

and is a functional of the argument function $z(x, y)$.

Other examples of functionals occurred in the previous chapter. Thus for fixed $K(x, y)$ the function

$$g(x) = \int K(x, y)h(y)\, dy$$

is a functional of $h(y)$ depending on x as a parameter,[2] and the integral form

$$\iint K(x, y)\varphi(x)\varphi(y)\, dx\, dy$$

is a functional of $\varphi(x)$. In this chapter we shall be concerned principally with functionals which are integrals of known functions of the argument function, its derivatives, and independent variables, such as the arc length of a curve.

A region of definition is needed for a finite number of variables; here we must define the *domain* or "*space*" *of admissible functions* from which the argument functions can be selected. This space may, for example, be the set of all functions which are continuous and have piecewise continuous first derivatives (compare the following section).

Although a functional cannot be expressed as a function of a finite number of variables, it may be regarded as a *function of infinitely many variables*. Suppose that the argument functions are expanded

[1] Throughout this book we shall employ the subscript notation for partial derivatives, i.e.

$$f_x = \partial f(x, y, \cdots)/\partial x; \qquad f_{xy} = \partial^2 f/\partial x \partial y, \text{ etc.}$$

[2] Often the words "function of a function" or "functional transform" are used for functionals which themselves depend on variables.

in power series or Fourier series; the functional then depends on the expansion coefficients, which constitute the infinitely many variables. Of course the domain of these variables must be restricted so as to conform to the conditions imposed on the argument functions.

3. Typical Problems of the Calculus of Variations. The calculus of variations is concerned with the problem of determining maxima or minima or, in general, stationary values[1] of functionals by seeking that argument function in the given domain of admissible functions for which the functional assumes the stationary value or extremum in question. Ordinary extremum problems of the differential calculus do not usually involve absolute maxima or minima but only *an extremum* with respect to values of the function in the vicinity of the extremum point; in analogy, we shall seek here the extremum of the functional only with respect to argument functions in a certain neighborhood of the extremal[2] argument function. For this purpose it is necessary to define the concept of the *neighborhood of a function* $f(x, y, \cdots)$. If h is a positive quantity, a function $f_1(x, y, \cdots)$ is said to lie in the neighborhood (h) of the function $f(x, y, \cdots)$ if $|f - f_1| < h$ throughout the region of definition.[3]

We are now in a position to formulate the *fundamental problem of the calculus of variations*: In a given domain of admissible argument functions, find an argument function (or functions) of a given functional for which the latter is an extremum with respect to all argument functions of the domain lying in a sufficiently small neighborhood (h) of the extremal argument function. If the functional to be made an extremum depends explicitly on variable parameters $x, y, \cdots$ in addition to the argument functions, i.e. if the functional is not a number but a function of these parameters, then we must determine these parameters as well as the argument functions, in order to obtain the extremum. We shall now illustrate the problems of the variational calculus by a number of simple examples.

[1] In §3, 1 the "stationary value of a functional" will be defined precisely.

[2] An extremal function is a function which makes a given functional stationary.

[3] For certain problems it is convenient to refine the concept of neighborhood in successive steps. The function $f_1(x, y, \cdots)$ is said to be in the neighborhood of first order (h) of $f(x, y, \cdots)$ if the relations $|f_x - f_{1x}| < h$, $|f_y - f_{1y}| < h$, $\cdots$ are satisfied in addition to $|f - f_1| < h$. More generally, we speak of a neighborhood of order $n + 1$ if these inequalities hold for f_1 and all its derivatives up to and including the n-th order.

(a) *Geodesic Curves.* On a given surface, find the shortest curve between two given points. If the surface is given by the parametric representation $x = x(u, v)$, $y = y(u, v)$, $z = z(u, v)$ in rectangular coordinates x, y, z, and if we employ the customary notatious $e = x_u^2 + y_u^2 + z_u^2$, $f = x_u x_v + y_u y_v + z_u z_v$, $g = x_v^2 + y_v^2 + z_v^2$ then the length L of a curve on the surface defined by the equation $v = v(u)$ between u_0 and u_1 is given by the integral

$$L = \int_{u_0}^{u_1} \sqrt{e + 2fv' + gv'^2}\, du.$$

The problem is thus to find functions $v(u)$ which yield extrema of this integral.

(b) *Light Rays; Brachistochrone.* According to Fermat's principle (page 165) the path of a light ray in an inhomogeneous two-dimensional medium in which the velocity of light is $c(x, y)$ solves the variational problem

$$T = \int_{x_0}^{x_1} \frac{\sqrt{1 + y'^2}}{c(x, y)}\, dx = \min.$$

In this problem, as in the previous one, all continuous curves are admissible which have piecewise continuous derivatives and join the given end points of the path. Closely related to the problem of the light ray is the brachistochrone problem, with which Jakob Bernoulli in 1696 gave the initial impetus to the development of the calculus of variations. Two points $A(x_0, 0)$ and $B(x_1, y_1)$ with $y_1 > 0$ are to be connected by a curve along which a frictionless mass point moves in the shortest possible time from A to B under gravity acting in the y-direction. The initial velocity of the mass point is zero. After falling a distance y the point has the velocity $\sqrt{2gy}$, according to elementary mechanics, where g is the acceleration due to gravity. It follows that the time of transit is the integral

$$T = \int_{x_0}^{x_1} \sqrt{\frac{1 + y'^2}{2gy}}\, dx. \tag{2}$$

The class of admissible functions consists of all the positive functions $y(x)$ with continuous second derivatives for which $y(x_0) = 0$ and $y(x_1) = y_1$.

(c) *Minimal Surface of Revolution.* Suppose that the curve $y = y(x) \geq 0$ is rotated about the x-axis. The resulting surface, bounded by planes $x = x_0$ and $x = x_1$, has the area $F = 2\pi \int_{x_0}^{x_1} y\sqrt{1 + y'^2}\, dx$.

The curve $y = y(x)$ which leads to the smallest surface of revolution is thus characterized by the variational problem

$$\int_{x_0}^{x_1} y\sqrt{1 + y'^2}\, dx = \min.$$

(d) *Isoperimetric Problem.* The original formulation of this problem is: Find a closed curve of given circumference and greatest area. Assuming the curve to be convex and divided into two parts of equal area by the x-axis (see subsection 1(d)) we arrive at the following problem: The integral

$$\int_0^{\xi} y(x)\, dx$$

is to be made a maximum by a suitable choice of ξ and $y(x)$, while the intergal

$$\int_0^{\xi} \sqrt{1 + y'^2}\, dx = l$$

has a given value. Here $y(x)$ may be any function which is continuous and has a piecewise continuous first derivative in the interval $0 \leq x \leq \xi$ and for which $y(0) = y(\xi) = 0$.

An analogous problem may be formulated with the upper limit ξ fixed.

This problem, known as the special isoperimetric problem, may be reduced to a simpler type of variational problem by introducing as the independent variable the arc length $s = \int_0^x \sqrt{1 + y'^2}\, dx$, ranging through the interval $0 \leq s \leq l$. Since $ds^2 = dx^2 + dy^2$, the problem now becomes: Find the function $y(s)$ which maximizes the integral

$$\int_0^l y\sqrt{1 - (dy/ds)^2}\, ds,$$

where $y(s)$ is a continuous function of s with a piecewise continuous derivative. After determining $y(s)$ we find

$$x(s) = \int_0^s \sqrt{1 - \left(\frac{dy}{ds}\right)^2}\, ds \tag{3}$$

and thus obtain a parametric representation of the desired curve. (Compare also Hurwitz's solution of the isoperimetric problem, Ch. II, §10, 1.)

In general, all problems in which one integral expression is to be made an extremum while another has a given value are termed *isoperimetric problems*. An example is given by the *problem of the catenary*: Determine the location of a uniform string of given length with fixed end points under the influence of gravity. Suppose that gravity acts in the negative y-direction. Since the equilibrium position demands that the center of gravity be as low as possible, we arrive at the following variational problem: Find the function $y(x)$ for which

$$\int_{x_0}^{x_1} y\sqrt{1+y'^2}\,dx$$

is as small as possible while

$$l = \int_{x_0}^{x_1} \sqrt{1+y'^2}\,dx$$

has a given value, and the boundary values $y(x_0) = y_0$, $y(x_1) = y_1$ are given.

Another problem is to solve

$$\int_{x_0}^{x_1} (y'')^2\,dx = \min.$$

with the subsidiary condition

$$\int_{x_0}^{x_1} y^2\,dx = 1,$$

where the function $y(x)$ is to vanish at the end points of the interval and is to be continuous everywhere along with its derivatives up to the second order. Another very important isoperimetric problem is: Find a function u of x and y, continuous along with its first derivatives in the region G, which satisfies $\iint_G u^2\,dx\,dy = 1$ and for which

$$\iint_G (u_x^2 + u_y^2)\,dx\,dy + \int_\Gamma \sigma u^2\,ds \tag{4}$$

is as small as possible[1] (here Γ is the boundary of the region G and σ is a fixed function of the arc length s of Γ).

The minimum problem of Chapter III for the eigenfunctions

[1] Cf. §3, 2.

of a symmetric kernel is another example of an isoperimetric problem.

It goes without saying that in all these problems the space of admissible argument functions must be specified by continuity conditions such that all the functionals occurring in the problem are meaningful.

4. Characteristic Difficulties of the Calculus of Variations. In the theory of ordinary maxima and minima the existence of a solution is ensured by the fundamental theorem of Weierstrass. In contrast, the characteristic difficulty of the calculus of variations is that problems which can be meaningfully formulated may not have solutions—because it is not in general possible to choose the domain of admissible functions as a "compact set" in which a principle of points of accumulation is valid. A simple geometric example is the following: Two points on the x-axis are to be connected by the shortest possible line of continuous curvature which is perpendicular to the x-axis at the end points. This problem has no solution. For, the length of such a line is always greater than that of the straight line connecting the two points, but it may approximate this length as closely as desired. Thus there exists a greatest lower bound but no minimum for admissible curves.

Another example of an insoluble variational problem is the following: Minimize the integral

$$\int_{-1}^{1} x^4 y'^2 \, dx \tag{5}$$

by means of a continuous function $y(x)$ which has a piecewise continuous derivative and for which $y(-1) = -1$, $y(1) = 1$. It is easily seen that the integral can be made arbitrarily small by suitable functions (i.e. $y = -1$ for $x < -\epsilon$, $y = x/\epsilon$ for $|x| \leq \epsilon$, $y = 1$ for $x > \epsilon$), but does not vanish for any admissible function.

Thus, *in the calculus of variations the existence of an extremum in a particular problem cannot be taken for granted.* A special existence proof is needed for the solution of each problem or class of problems. As we shall see later, this fact gives rise to essential difficulties in many of the problems of the calculus of variations. However, in this chapter we shall deal primarily with the formulation of *necessary conditions* for the attainment of an extremum, while the question whether an extremum is actually attained when the conditions are satisfied may be left open.

Before formulating these necessary conditions we shall consider possible methods for the direct solution of variational problems.

§2. *Direct Solution*[1]

Direct and complete solutions of variational problems can sometimes be obtained by a general procedure consisting of two steps: first one formulates a suitable approximate ordinary extremum problem in which only a finite number n of parameters is to be determined and then one passes to the limit $n \to \infty$ in the solution of the approximate problem.

1. The Isoperimetric Problem. One example is the isoperimetric problem of §1, 3(d): find a closed curve K with perimeter $2l$ and maximum area; the curve is to be piecewise smooth, i.e. it is to have a continuous tangent except at a finite number of corners. If we assume that the solution is given by the curve K we may conclude that K is a circle. For, just as in §1, 1(d), it may be proved that K is convex and that every secant AB which divides K into two parts with equal circumference also divides it into two parts of equal area; furthermore, for every point P on K the angle APB must be a right angle, since otherwise, by the construction of §1, 1(d), one could obtain a curve K' with the same perimeter but greater area. But this argument is based on the tacit assumption that the problem actually possesses a solution, and this assumption requires proof; therefore we shall solve the problem by a different method which furnishes the necessary existence proof at the same time. We consider the set of values of the areas of admissible curves. Since these numbers all lie below the bound $l^2\pi$ (the curve can certainly be placed inside a circle of radius l), this set of numbers has, by the elementary rules of analysis, a least upper bound M, above which there is no number of the set, but such that numbers of the set occur in any neighborhood (ϵ) of M for arbitrarily small ϵ. In other words, there exists a "maximizing sequence" of admissible curves $K_1, K_2, \cdots$ such that the area F_n of K_n converges to M with increasing n. Now every curve K_n can be approximated as exactly as desired by an inscribed polygon Π_n with sufficiently many sides such that area and perimeter of Π_n differ arbitrarily little from those of K_n. We may, without destroying its approximating character, dilate each polygon in such a way that its

[1] Direct methods of the calculus of variations will be treated more fully in the second volume.

perimeter becomes exactly equal to $2l$. Thus the maximizing sequence $K_1, K_2, \cdots$ may be replaced by a maximizing sequence of polygons $\Pi_1', \Pi_2', \cdots$. The number of sides of these polygons may be assumed to be even, since a $(2m-1)$-gon may be regarded as a $2m$-gon with two adjacent sides forming part of the same straight line. We know from §1, 1(d) that of all $2m$-gons of perimeter $2l$ the regular $2m$-gon has the greatest area. We therefore obtain an even better maximizing sequence of our problem by replacing each polygon Π_n' by a corresponding regular polygon. But as the number of sides increases these polygons converge to the circle of perimeter $2l$, and since the areas of the polygons converge to M the circle has the area M and is actually the solution of our variational problem.

2. The Rayleigh-Ritz Method. Minimizing Sequences. The above considerations are based on the following general idea: We consider any variational problem of the form $D[\varphi] = \text{min.}$, where the functional $D[\varphi]$ is an integral of a given expression containing the function φ and its derivatives up to the k-th order, and where the region of integration and the domain of admissible functions φ are given. It does not matter whether $D[\varphi]$ is a simple or a multiple integral. We assume that the set of values of $D[\varphi]$ for the admissible argument functions φ possesses a greatest lower bound d (whether d is a minimum which is actually attained for a function $\varphi = u$ remains an open question). Then there exist sequences $\varphi_1, \varphi_2, \cdots$ of admissible functions such that $\lim_{n\to\infty} D[\varphi_n] = d$, while the relation $D[\varphi] \geq d$ holds for every admissible function φ. Such sequences of functions will be called *minimizing sequences* of the variational problem. *A direct solution of the variational problem always consists in the construction of minimizing sequences and the attempt to secure the solution by a limiting process based on this sequence.*

The method which W. Ritz[1] used with great success, especially for numerical solutions, consists in the following steps: We start with a fixed complete system of "coordinate" functions $\omega_1, \omega_2, \cdots$, defined in the region of integration, which has the property[2] that all linear combinations $\varphi_n = c_1\omega_1 + c_2\omega_2 + \cdots + c_n\omega_n$ of a finite number

[1] W. Ritz, Über eine neue Methode zur Lösung gewisser Variationsprobleme der mathematischen Physik, Journ. f. d. reine u. angew. Math., Vol. 135, 1909, pp. 1–61; Gesammelte Werke, pp. 192–250, Gauthier-Villars, Paris, 1911. Even before Ritz, similar ideas were successfully employed by Lord Rayleigh.

[2] The question of the existence of such functions will be discussed more fully in the second volume.

of the functions are admissible comparison functions for the problem, and such that for every admissible function φ there is a suitable combination φ_n of this kind for which $D[\varphi]$ differs arbitrarily little from $D[\varphi_n]$. Then there exist minimizing sequences $\varphi_1, \varphi_2, \cdots$ in which φ_n is a linear combination $c_1\omega_1 + c_2\omega_2 + \cdots + c_n\omega_n$ of the functions $\omega_1, \omega_2, \cdots, \omega_n$. We therefore obtain an even better minimizing sequence if for every n we determine the function φ_n, i.e. the parameters $c_1, c_2, \cdots, c_n$, by the requirement that $D[\varphi_n] = d_n$ be a minimum. This requirement represents an ordinary minimum problem for $D[\varphi_n]$ as a function of the n parameters $c_1, c_2, \cdots, c_n$ and can always be fulfilled, according to Weierstrass's theorem (provided $D[\varphi_n]$ is a continuously differentiable function of $c_1, c_2, \cdots c_n$; we shall assume here that this is the case). The values c_i are determined by the n simultaneous equations $\partial D[\varphi_n]/\partial c_i = 0$ $(i = 1, 2, \cdots, n)$. We may now expect the resulting minimizing sequence to converge to the desired solution. Unfortunately things are not so simple, as we shall see in §4. In general we can only state that the values $D[\varphi_n] = d_n$ obtained in this way converge to the desired greatest lower bound or minimum. Whether the minimizing sequence itself converges to the solution is a difficult theoretical question which has to be investigated separately. We shall return to this question on several subsequent occasions.

In some cases this method may prove useful for numerical calculations even though its convergence is unproved. Its success in any particular case will depend on the proper choice of *coordinate functions* ω_i. The method will be illustrated by examples in the next subsection.

3. Other Direct Methods. Method of Finite Differences. Infinitely Many Variables. One can often obtain minimizing sequences by extending the domain of admissible functions, for example by admitting not only continuously differentiable functions, but also continuous functions with piecewise continuous derivatives. We shall consider the problem of the minimum of a simple integral of the form $D[y] = \int_{x_0}^{x_1} F(x, y, y')\,dx$. If $y = \varphi_1(x)$, $y = \varphi_2(x)$, $\cdots$ is a minimizing sequence, and if $F(x, y, y')$ satisfies the necessary continuity requirements, then the curve given by $y = \varphi_n(x)$ can certainly be approximated by a polygonal arc $y = p_n(x)$ in such a way that the integral $D[\varphi_n]$ differs arbitrarily little from $D[p_n]$. It is therefore

possible to construct minimizing sequences consisting of piecewise linear functions, for which the difference quotient and the differential quotient are identical in each subinterval. Thus if we divide the interval of integration by m points into equal parts of length Δx, and if we confine ourselves to functions which are linear in each subinterval, the variational problem goes over into, or may be approximately replaced by, the ordinary minimum problem

$$\sum_{i=0}^{m} F\left(x_i, y_i, \frac{y_{i+1} - y_i}{\Delta x}\right) \Delta x = \min.$$

for the values $y_0, y_1, \cdots, y_{m+1}$ of the function at the ends of the subintervals. The resulting functions for $m = 0, 1, 2, \cdots$ once again form a minimizing sequence.[1]

This method may be regarded as a special case of the Ritz method, with suitable piecewise linear coordinate functions.

Under what circumstances the solution of the finite difference problem converges to the solution of the minimum problem will be discussed in the second volume.

It is possible to proceed in a similar manner when the integrand contains derivatives of higher order, say the second. The second differential quotients will then be replaced in the approximating problem by the second difference quotients $(y_{i+2} - 2y_{i+1} + y_i)/(\Delta x)^2$.

Our variational problems may also be regarded as problems in the theory of functions of infinitely many variables. For example, in Hurwitz's solution of the isoperimetric problem (Ch. II, p. 97) the variables in question were the Fourier coefficients and the analytic expression for $L^2 - 4\pi F$ exhibits the solution. The Ritz method also admits of this interpretation if we may assume the function to be developed in a Fourier series $c_1\omega_1 + c_2\omega_2 + \cdots$ and regard the method as a prescription for obtaining the infinitely many coefficients $c_1, c_2, \cdots$. The difficult question of convergence would have to be investigated separately.

We shall now illustrate these general considerations by some examples.

[1] The method described here is essentially the method which led Euler to the "Euler differential equations" in his work: Methodus inveniendi lineas curvas maximi minimive proprietate gaudentes, M. Bousquet, Lausanne and Geneva, 1744.

(a) The integral

$$D[\varphi] = \iint_R (\varphi_x^2 + \varphi_y^2)\, dx\, dy, \tag{6}$$

which extends over the rectangle R given by $0 \leq x \leq a$, $0 \leq y \leq b$, is to be minimized; here the admissible comparison functions are all piecewise smooth[1] in R, vanish on its boundary, and satisfy the subsidiary condition

$$H[\varphi] = \iint_R \varphi^2\, dx\, dy = 1. \tag{7}$$

We suppose that the function φ is expanded in a Fourier series $\varphi = \sum_{m,n=1}^{\infty} c_{mn} \sin(m\pi x/a) \sin(n\pi y/b)$; this is certainly possible according to Chapter II. The problem is to determine the infinitely many parameters c_{mn} from the minimum requirement. Since the functions φ_x and φ_y are piecewise continuous we may apply the completeness relation of the trigonometric functions to these functions with the expansion coefficients $(\pi m/a)c_{mn}$, $(\pi n/b)c_{mn}$, obtaining for the two integrals the expressions

$$D = \pi^2 \frac{ab}{4} \sum_{m,n=1}^{\infty} c_{mn}^2 \left(\frac{m^2}{a^2} + \frac{n^2}{b^2}\right); \qquad H = \frac{ab}{4} \sum_{m,n=1}^{\infty} c_{mn}^2 \tag{8}$$

in the parameters c_{mn}. Because of the condition $H = 1$ it is now evident that the solution of the problem is given by $c_{mn} = 0$ except for the coefficient c_{11}, which must equal $2/\sqrt{ab}$. Thus our variational problem is solved by the function

$$u = \frac{2}{\sqrt{ab}} \sin\frac{\pi x}{a} \sin\frac{\pi y}{b},$$

and the value of the minimum is

$$d = \pi^2\left(\frac{1}{a^2} + \frac{1}{b^2}\right).$$

This may be expressed by relation

$$D[\varphi] \geq \pi^2\left(\frac{1}{a^2} + \frac{1}{b^2}\right) H[\varphi] \tag{9}$$

[1] See the definition given in Ch. II, p. 48.

for every piecewise smooth function vanishing on the boundary of the rectangle, because this relation is equivalent to $D[\psi] \geq d$ for the normalized function $\psi = \varphi/\sqrt{H[\varphi]}$.

(b) *Dirichlet's Problem*[1] *for the Circle.* The integral

$$D[\varphi] = \iint_K (\varphi_x^2 + \varphi_y^2)\, dx\, dy,$$

extended over the circle K $(x^2 + y^2 \leq 1)$ in the x, y-plane, is to be made a minimum if the admissible comparison functions are smooth in K and assume given boundary values on the boundary of K. We introduce polar coordinates r, θ in K, transforming the integral into

$$D[\varphi] = \int_0^{2\pi} \int_0^1 (\varphi_r^2 + \frac{1}{r^2}\varphi_\theta^2) r\, dr\, d\theta.$$

The boundary values may now be defined by a Fourier series $f(\theta) = \frac{1}{2}a_0 + \sum_{n=1}^{\infty}(a_n \cos n\theta + b_n \sin n\theta)$. Suppose further that this boundary function f has a smooth derivative; according to Ch. II, §5, 3, this means that $n^2|\, a_n\,|$ and $n^2|\, b_n\,|$ are bounded. We may now assume the function φ written in the form

$$\varphi = \tfrac{1}{2} f_0(r) + \sum_{n=1}^{\infty} [f_n(r) \cos n\theta + g_n(r) \sin n\theta],$$

where $f_n(r)$, $g_n(r)$ must satisfy the relations $f_n(1) = a_n$, $g_n(1) = b_n$. Because of the completeness relation for the trigonometric functions we obtain the equation

$$D[\varphi] = \pi \int_0^1 [f_0'(r)]^2 r\, dr + \pi \sum_{n=1}^{\infty} \int_0^1 \left[f_n'^2(r) + \frac{n^2}{r^2} f_n^2(r) \right] r\, dr$$
$$+ \pi \sum_{n=1}^{\infty} \int_0^1 \left[g_n'^2(r) + \frac{n^2}{r^2} g_n^2(r) \right] r\, dr.$$

Thus in order to solve the original minimum problem we treat the series of minimum problems

$$\int_0^1 \left(f_n'^2 + \frac{n^2}{r^2} f_n^2 \right) r\, dr = \text{min.}, \qquad \int_0^1 \left(g_n'^2 + \frac{n^2}{r^2} g_n^2 \right) r\, dr = \text{min.}$$
$$(n = 0, 1, 2, \cdots),$$

[1] This nomenclature has become customary since Riemann's time even though it does not correspond to the historical facts.

in which f_n and g_n are smooth functions that take on the values a_n and b_n, respectively, at $r = 1$. In view of Weierstrass's approximation theorem the functions $1, r, r^2, \cdots$ certainly satisfy the conditions for the coordinate functions of the Ritz method for these minimum problems. Let us take for f_n (or g_n) a polynomial $c_{n,0} + c_{n,1}r + \cdots + c_{n,m}r^m$ with $m \geq n$ and with $c_{n,0} + c_{n,1} + \cdots + c_{n,m} = a_n$ (or b_n). The reader may easily verify that the solutions, independent of m, are $f_n = a_n r^n$ (or $g_n = b_n r^n$). Thus all the functions of the resulting minimizing sequence are equal to each other, and therefore equal to the solution of the variational problem.

The solution f_n or g_n of the variational problem may also be found directly, without employing the Ritz procedure, in the following way: For $n = 0$ we have to minimize $\int_0^1 f_0'^2 r\, dr$; this is evidently accomplished with $f_0' = 0$, $f_0 = \text{const.} = a_0$. For $n > 0$ we must have $f_n(0) = g_n(0) = 0$; otherwise the second part of the integral would be infinite, since f_n is differentiable and may therefore be written in the form $f_n(0) + rh_n(r)$ with a continuous function $h_n(r)$. We now write the first integral in the form

$$\int_0^1 \left(f_n' - \frac{n}{r} f_n\right)^2 r\, dr + 2n \int_0^1 f_n f_n'\, dr = \int_0^1 \left(f_n' - \frac{n}{r} f_n\right)^2 r\, dr + n f_n^2(1).$$

The value of $f_n(1) = a_n$ is fixed, and we see immediately that we obtain a minimum value, namely $nf_n^2(1) = na_n^2$, if we set $f_n' - nf_n/r = 0$, obtaining $f_n = c_n r^n$; since $f_n(1) = a_n$ we have $c_n = a_n$ or $f_n = a_n r^n$. Corresponding results are obtained for $g_n(r)$. Thus the solution of the original minimum problem is

$$u(r, \theta) = \tfrac{1}{2} a_0 + \sum_{n=1}^{\infty} r^n (a_n \cos n\theta + b_n \sin n\theta) \tag{10}$$

and the minimum value is

$$\pi \sum_{n=1}^{\infty} n(a_n^2 + b_n^2).$$

It should be emphasized that the minimum problem solved here may cease to be meaningful if less stringent assumptions are made

concerning the boundary function, for example if continuity alone is required. Thus, if we take the continuous boundary function defined by the uniformly convergent series $\rho(\theta) = \sum_{n=1}^{\infty} \frac{1}{n^2} \cos(n!\theta)$, the sum $\pi \sum_{m=1}^{\infty} m a_m^2 = \pi \sum_{n=1}^{\infty} \frac{1}{n^4} (n!)$ is infinite. In this case it can easily be shown that there exists no admissible comparison function with finite $D[\varphi]$.

(c) Let $g(x, y)$ be a smooth function in the rectangle R defined by $0 \leq x \leq a$, $0 \leq y \leq b$. We seek a function φ, continuous in the rectangle along with its first derivatives and vanishing on the boundary, which minimizes the integral

$$J[\varphi] = \iint_R (\varphi_x^2 + \varphi_y^2 - 2\varphi g)\, dx\, dy. \tag{11}$$

In the interior of R, let us write

$$g(x, y) = \sum_{m,n=1}^{\infty} a_{mn} \sin m \frac{\pi x}{a} \sin n \frac{\pi y}{b}$$

and

$$\varphi = \sum_{m,n=1}^{\infty} c_{mn} \sin m \frac{\pi x}{a} \sin n \frac{\pi y}{b},$$

where the parameters c_{mn} are to be determined. Because of the completeness relation for the trigonometric functions the variational problem then goes over into the problem of determining the quantities c_{mn} such that

$$\frac{4}{ab} J[\varphi] = \pi^2 \sum_{m,n=1}^{\infty} \left(\frac{m^2}{a^2} + \frac{n^2}{b^2}\right) c_{mn}^2 - 2 \sum_{m,n=1}^{\infty} a_{mn} c_{mn}$$

becomes as small as possible. We see immediately that the minimum is furnished by

$$\begin{aligned} c_{mn} &= \frac{a_{mn}}{\pi^2 (m^2/a^2 + n^2/b^2)}, \\ u(x, y) &= \frac{1}{\pi^2} \sum_{m,n=1}^{\infty} \frac{a_{mn}}{m^2/a^2 + n^2/b^2} \sin m \frac{\pi x}{a} \sin n \frac{\pi y}{b}. \end{aligned} \tag{12}$$

This function does indeed satisfy all requirements, since the series, as well as the series obtained by term-by-term differentiation, con-

verges uniformly. To see this we note that the absolutely convergent series

$$\sum \frac{|a_{mn}|}{m^2/a^2 + n^2/b^2}, \qquad \sum \frac{m|a_{mn}|}{m^2/a^2 + n^2/b^2}, \qquad \sum \frac{n|a_{mn}|}{m^2/a^2 + n^2/b^2}$$

represent majorants. We shall see subsequently (page 192) that the function u satisfies the differential equation $u_{xx} + u_{yy} = g(x, y)$; compare examples (a) and (b).

4. General Remarks on Direct Methods of the Calculus of Variations. We have already noted that the main difficulty in justifying the direct methods of the calculus of variations is that minimizing sequences do not necessarily converge to a limit function, even if the existence of a solution is not in question.

A simple example is given by the variational problem of minimal surfaces, in which the integral

$$\iint_G \sqrt{1 + z_x^2 + z_y^2}\, dx\, dy$$

is to be minimized; here we admit all surfaces $z = z(x, y)$ with piecewise continuous derivatives which pass through a given space curve whose projection is the boundary of G. If, in particular, we take for this curve a curve in the x, y-plane, for example a circle of unit area, the minimum is given by the function $z = 0$, i.e. by the x, y-plane itself. Every sequence of surfaces which passes through the periphery of the circle and the areas of which converge to 1 is a minimizing sequence. Now we may at once construct admissible comparison surfaces whose areas are arbitrarily close to 1 and for which $z(x, y)$ is arbitrarily large at isolated points. For example, consider a circular cone of height 1 and base radius ϵ placed perpendicularly on the plane $z = 0$. We now take as a comparison surface the surface consisting of this cone and of the portion of the original circle outside the base of the cone. A minimizing sequence of such surfaces does not, for $\epsilon \to 0$, converge to the solution. It is even possible, as is easily seen, to construct minimizing sequences in which the points of divergence are everywhere dense in the circle.

Another example is given by the Dirichlet problem of minimizing the integral $D[\varphi] = \iint_G (\varphi_x^2 + \varphi_y^2)\, dx\, dy$ if we admit for comparison all functions continuous and piecewise smooth in G which vanish on the boundary. Evidently the unique solution of the problem is $\varphi = 0$,

since every other admissible function leads to a positive value of the integral. Introducing polar coordinates r, θ about an arbitrary point P of G, we write the integral in the form $\iint_G (\varphi_r^2 + \varphi_\theta^2/r^2) r \, dr \, d\theta$. We now consider a circle $r \leq a$ about P, lying entirely in the interior of G, with radius $a < 1$. Let us set $\varphi = 0$ outside this circle, $\varphi = \log (r/a)/\log a$ in the circular ring between $r = a$ and $r = a^2$, and $\varphi = \log a/\log a = 1$ in the circle $r \leq a^2$. By definition, φ is an admissible comparison function; therefore the integral is equal to

$$\frac{2\pi}{(\log a)^2} \int_{a^2}^{a} \frac{1}{r^2} \, r \, dr = -\frac{2\pi}{\log a}.$$

If we now let a assume a sequence of values $a_1, a_2, \cdots$ which tend to zero and consider the corresponding functions $\varphi_1, \varphi_2, \cdots$, we see that $D[\varphi_n]$ converges to zero; thus these functions form a minimizing sequence of the problem. But at the point P all the functions have the value 1 and therefore do not converge toward the solution $\varphi = 0$ of the problem.

In the case of the variational problem $\int_0^1 y'^2 \, dx = \text{min.}$, where $y(x)$ is a continuous piecewise smooth function of x vanishing at the end points, all minimizing sequences must indeed converge to the function $y = 0$, as is easily seen. But the derivatives of the functions of the minimizing sequence do not necessarily converge to zero, as is shown by the example of the sequence $y_n = x$ for $x < \epsilon_n$, $y_n = 2\epsilon_n - x$ for $\epsilon_n \leq x \leq 2\epsilon_n$, $y_n = 0$ for $x > 2\epsilon_n$, $(\lim_{n\to\infty} \epsilon_n = 0)$.

In Volume II we shall see how these difficulties may be overcome; indeed, the direct methods of the variational calculus have led to important developments in analysis.

The following sections deal with classical "indirect" methods based on the reduction of variational problems to differential equations. These methods, which have been emphasized by mathematicians since Euler and Lagrange, attack minimum problems by means of a general variational formalism which in itself has become an important tool of analysis.

§3. *The Euler Equations*

The differential equations of a variational problem, first derived by Euler, represent necessary but by no means always sufficient conditions which a function must satisfy if it is to furnish the extremum

of a given integral. We obtain these Euler equations by reducing the variational problem to a problem in the differential calculus. Once and for all we assume that all functions and their derivatives which occur explicitly in the problems are continuous unless the contrary is explicitly stated.

1. "Simplest Problem" of the Variational Calculus. Let us first consider the simplest problem of the calculus of variations, namely that of determining the minimum of the integral

$$J[y] = \int_{x_0}^{x_1} F(x, y, y')\, dx, \tag{13}$$

where the values x_0, x_1, $y(x_0)$, $y(x_1)$ are given. The function F is to be twice continuously differentiable with respect to its three arguments x, y, y'. The second derivative y'' of the function y is also assumed continuous. Suppose $y = y(x) = f(x)$ is the desired extremal function yielding the minimum, i.e. suppose that, in a sufficiently small neighborhood (h) of the function $f(x)$, the integral $J[y]$ is smallest when $y = f(x)$. Now consider a function $\eta(x)$ which is defined in the interval $x_0 \leq x \leq x_1$, possesses a continuous second derivative, and vanishes at the end points, but is otherwise arbitrary. We construct the function $\bar{y} = y + \epsilon\eta(x) = y + \delta y$, where ϵ is a parameter. The quantity $\delta y = \epsilon\eta(x)$ is known as the *variation* of the function $y = f(x)$. If the parameter ϵ has a sufficiently small absolute value all the varied functions $\bar{y}$ lie in an arbitrarily small neighborhood of the extremal function $y = f(x)$. Therefore the integral $J[\bar{y}] = J[y + \epsilon\eta]$, which may be regarded as a function $\Phi(\epsilon)$ of ϵ, must have a minimum at $\epsilon = 0$ relative to all values of ϵ in a sufficiently small neighborhood of 0, and therefore $\Phi'(0) = 0$. Now if we differentiate the integral $\Phi(\epsilon) = \int_{x_0}^{x_1} F(x, y + \epsilon\eta, y' + \epsilon\eta')\, dx$ with respect to ϵ under the integral sign (this is permissible) we obtain as a necessary condition the equation

$$\Phi'(0) = \int_{x_0}^{x_1} (F_y\eta + F_{y'}\eta')\, dx = 0,$$

which must hold for all functions $\eta(x)$ which satisfy the above requirements. We transform the second part of the integral by partial integration, noting that $\eta(x_0) = \eta(x_1) = 0$, and obtain the equation

$$\int_{x_0}^{x_1} \eta(F_y - \frac{d}{dx} F_{y'}) \, dx = 0,$$

valid for every one of our functions η.

This equation immediately leads to the desired differential equation in virtue of the following *fundamental lemma of the calculus of variations*: If the relation $\int_{x_0}^{x_1} \eta(x)\varphi(x) \, dx = 0$, with $\varphi(x)$ a continuous function of x, holds for all functions $\eta(x)$ which vanish on the boundary and are continuous together with their first two derivatives, it follows that $\varphi(x) = 0$ identically. This lemma, which holds equally well for multiple integrals, is easily proved indirectly. Let us suppose that $\varphi(x)$ is different from zero, say positive, at $x = \xi$. Then there must exist a neighborhood G, given by $\xi_0 < x < \xi_1$, in which $\varphi(x)$ is positive. We take the function $\eta(x) = (x - \xi_0)^4(x - \xi_1)^4$ in G and $\eta(x) = 0$ outside this interval. Then we have $\int_{x_0}^{x_1} \eta\varphi \, dx > 0$ in contradiction to the hypothesis. The assertion $\varphi = 0$ remains valid if we require that the first k derivatives of the function η be continuous; in this case we simply set $\eta = (x - \xi_0)^{2l}(x - \xi_1)^{2l}$ with $2l > k$.

From the fundamental lemma it follows immediately that the function $dF_{y'}/dx - F_y$, which we shall henceforth denote by $-[F]_y$, vanishes identically in x, i.e. that the function $y(x)$ satisfies the differential equation

$$-[F]_y = \frac{d}{dx} F_{y'} - F_y = 0 \tag{14}$$

or, written out in detail,

$$y''F_{y'y'} + y'F_{y'y} + F_{y'x} - F_y = 0. \tag{14'}$$

This is the fundamental *differential equation of Euler*, which occurs time and again throughout analysis and its applications. *Its validity is a necessary condition for the existence of an extremum.* It is a differential equation of the second order, in the general solution of which there occur two arbitrary constants, the number generally required in order to satisfy the boundary conditions. We now define: *every* solution of Euler's differential equation is an *extremal*[1] of the minimum problem. The differential expression $[F]_y$ is called the

[1] Cf. definition in footnote 2 on page 169.

variational derivative of F with respect to y. Its role here is analogous to that of the differential quotient or of the gradient in ordinary minimum problems.

If, as is customary in the theory of differential equations, we want to solve the Euler differential equation for the highest derivative, we have to postulate

$$F_{y'y'} \neq 0.$$

This inequality is known as the *Legendre condition*; it is of great importance in the problem of investigating whether an extremal actually gives an extremum (compare also §6, pages 214–216).

The essential idea in the above considerations is to imbed the extremal $y(x)$ in a family of functions $y(x) + \epsilon\eta(x)$ with the parameter ϵ. The fact that this parameter occurs linearly is not important. It is often convenient to imbed the function in a more general family of functions $y(x, \epsilon)$; the above considerations remain valid if we set

$$\eta(x) = \frac{\partial}{\partial \epsilon} y(x, \epsilon)\Big|_{\epsilon=0}.$$

We now introduce a useful terminology. We called the function $\epsilon\eta = \delta y$ the variation of $y(x)$; similarly the expression

$$\begin{aligned} \delta J &= \epsilon\,\Phi'(0) = \epsilon \int_{x_0}^{x_1} (\eta F_y + \eta' F_{y'})\, dx \\ &= \epsilon \int_{x_0}^{x_1} \left(F_y - \frac{d}{dx} F_{y'}\right) \eta\, dx + \epsilon F_{y'}\eta \,|^{x=x_1} - \epsilon F_{y'}\eta \,|^{x=x_0} \qquad (15) \\ &= \int_{x_0}^{x_1} [F]_y\, \delta y\, dx + F_{y'}\, \delta y\Big|_{x_0}^{x_1}, \end{aligned}$$

even if η does not vanish at the boundary, is called the *variation* or more precisely the *first variation* of the integral J. This is analogous to the notation of the differential calculus, where the expression $\epsilon f'(x) = df$ with an arbitrary parameter ϵ is called the differential of the function $f(x)$. Thus *a necessary condition for a minimum is the vanishing of the first variation* for all $y + \delta y$ which satisfy the boundary conditions.

In general, functions or, geometrically speaking, curves for which δJ vanishes, i.e. extremals, are sometimes called *stationary functions or curves*. This terminology indicates that we may be dealing with

circumstances other than actual extrema, just as in the corresponding problems of the differential calculus. There are, in fact, many cases in which one is primarily interested in the vanishing of the variation rather than in the question of an extremum. Problems in which one is merely interested in obtaining stationary values are likewise called variational problems.

The examples considered above (pages 170–173) lead to the following variational derivatives:

(a) $F = \sqrt{e + 2fv' + gv'^2}$,

$$\frac{d}{du}\frac{f + gv'}{\sqrt{e + 2fv' + gv'^2}} - \frac{e_v + 2f_v v' + g_v v'^2}{2\sqrt{e + 2fv' + gv'^2}} = 0;$$

(b) $F = \dfrac{\sqrt{1 + y'^2}}{\varphi(x, y)} = \psi(x, y)\sqrt{1 + y'^2}$,

$$\psi y'' = (\psi_y - \psi_x y')(1 + y'^2);$$

(c) $F = y\sqrt{1 + y'^2}$, $\quad yy'' = 1 + y'^2$ (a special case of (b));

(d) $F = y\sqrt{1 - y'^2}$, $\quad yy'' = y'^2 - 1$.

We shall return to the integration of these differential equations in §4.

2. Several Unknown Functions. A straightforward generalization of the problem discussed in subsection 1 is the problem of determining several functions $y(x)$, $z(x)$, $\cdots$ of x such that the integral

$$J = \int_{x_0}^{x_1} F(x, y, z, \cdots, y', z', \cdots)\, dx \tag{16}$$

is an extremum (or stationary value), where the values of the functions at the boundary points may again be given. Once again we introduce arbitrary functions $\eta(x)$, $\zeta(x)$, $\cdots$, vanishing on the boundary, and assume that the system of functions $y = y(x) = f(x)$, $z = z(x) = g(x)$, $\cdots$ yields an extremum. As above, this leads to the conclusion that the function

$$\Phi(\epsilon_1, \epsilon_2, \cdots) = \int_{x_0}^{x_1} F(x, y + \epsilon_1\eta, z + \epsilon_2\zeta, \cdots, y' + \epsilon_1\eta', z' + \epsilon_2\zeta', \cdots)\, dx$$

of the variables ϵ_1, ϵ_2, $\cdots$ must have an extremum at $\epsilon_1 = 0$,

$\epsilon_2 = 0, \cdots$. Thus we must have $(\partial\Phi/\partial\epsilon_1)_0 = 0$, $(\partial\Phi/\partial\epsilon_2)_0 = 0, \cdots$[1] or

$$\delta J = \epsilon_1\left(\frac{\partial\Phi}{\partial\epsilon_1}\right)_0 + \epsilon_2\left(\frac{\partial\Phi}{\partial\epsilon_2}\right)_0 + \cdots$$
$$= \int_{x_0}^{x_1} ((F_y\eta + F_{y'}\eta')\epsilon_1 + (F_z\zeta + F_{z'}\zeta')\epsilon_2 + \cdots)\,dx = 0.$$

This expression is called the first variation of J. Again it may be brought into the form

$$\delta J = \epsilon_1 F_{y'}\eta\Big|_{x_0}^{x_1} + \epsilon_2 F_{z'}\zeta\Big|_{x_0}^{x_1} + \epsilon_1\int_{x_0}^{x_1}\eta\left(F_y - \frac{d}{dx}F_{y'}\right)dx + \epsilon_2\int_{x_0}^{x_1}\zeta\left(F_z - \frac{d}{dx}F_{z'}\right)dx + \cdots; \tag{17}$$

in the present case the boundary terms vanish. Since we must have $\delta J = 0$ if one of the functions $\eta, \zeta, \cdots$ is chosen arbitrarily while all the others are zero, the arguments employed above show that the following Euler differential equations hold:

$$\begin{aligned} -[F]_y &= \frac{d}{dx}F_{y'} - F_y \\ &= F_{y'y'}y'' + F_{y'z'}z'' + \cdots + F_{y'y}y' + F_{y'z}z' + \cdots \\ &\qquad + F_{y'x} - F_y = 0, \\ -[F]_z &= \frac{d}{dx}F_{z'} - F_z \\ &= F_{z'y'}y'' + F_{z'z'}z'' + \cdots + F_{z'y}y' + F_{z'z}z' + \cdots \\ &\qquad + F_{z'x} - F_z = 0, \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \end{aligned} \tag{18}$$

Thus, as a necessary condition for the extremum or stationary character of the "space curve" $y = f(x)$, $z = g(x)$, $\cdots$, we obtain the system (18) of differential equations of the second order; the number of equations in this system equals the number of unknown functions $y, z, \cdots$ to be determined.

All considerations of subsection 1 remain valid. It may be noted

[1] The subscript 0 means that we set $\epsilon_1 = \epsilon_2 = \cdots = 0$.

that the vanishing of the first variation is a necessary condition not only in the case of an extremum, but for example, also in the case of an integral which is to be minimized with respect to variation of the function $y(x)$ and simultaneously maximized with respect to variation of the function $z(x)$.

Once again *every* curve which represents a solution of the system of differential equations is called an *extremal*.

A simple example of the Euler equations (18) is given by the problem of determining the shortest lines in ordinary Euclidean space, or more generally in a non-Euclidean space with the line element

$$ds^2 = g_{11}\,dx^2 + g_{22}\,dy^2 + g_{33}\,dz^2 + 2g_{12}\,dx\,dy + 2g_{13}\,dx\,dz + 2g_{23}\,dy\,dz.$$

In this case we have

$$F = (g_{11} + 2g_{12}y' + 2g_{13}z' + g_{22}y'^2 + 2g_{23}y'z' + g_{33}z'^2)^{1/2}$$

and we obtain for the "geodesic curves" of this space the two differential equations

$$\frac{d}{dx}\left(\frac{g_{12} + g_{22}y' + g_{23}z'}{F}\right) - \frac{1}{2F}\left(\frac{\partial g_{11}}{\partial y} + 2\frac{\partial g_{12}}{\partial y}y' + \cdots\right) = 0,$$

$$\frac{d}{dx}\left(\frac{g_{13} + g_{23}y' + g_{33}z'}{F}\right) - \frac{1}{2F}\left(\frac{\partial g_{11}}{\partial z} + 2\frac{\partial g_{12}}{\partial z}y' + \cdots\right) = 0.$$

In Euclidean space, where we have

$$ds^2 = dx^2 + dy^2 + dz^2, \qquad F = \sqrt{1 + y'^2 + z'^2},$$

these equations become

$$\frac{d}{dx}\frac{y'}{\sqrt{1 + y'^2 + z'^2}} = 0, \qquad \frac{d}{dx}\frac{z'}{\sqrt{1 + y'^2 + z'^2}} = 0$$

and are satisfied by all straight lines in space.

The propagation of light in a three-dimensional medium with light velocity $\varphi(x, y, z)$ is characterized by the variational problem

$$T = \int_{x_0}^{x_1} \frac{\sqrt{1 + y'^2 + z'^2}}{\varphi(x, y, z)}\,dx = \min.$$

More generally, we may suppose that the velocity of light depends on the direction of the ray as well, so that it is represented by an expression $\varphi(x, y, z, y', z')$. Then the problem of finding a light ray,

a problem in geometrical optics, is equivalent to our general problem, with

$$F = \frac{\sqrt{1 + y'^2 + z'^2}}{\varphi(x, y, z, y', z')}.$$

3. Higher Derivatives. Euler's differential equation is obtained in an analogous manner for the variational problem of finding stationary values of the integral

$$J = \int_{x_0}^{x_1} F(x, y, y', y'', \cdots, y^{(n)})\, dx, \tag{19}$$

where F is a given function of the arguments x, y, y', $\cdots$, $y^{(n)}$ and where all those functions are admitted for comparison which have continuous derivatives up to the $2n$-th order and for which the derivatives up to the $(n-1)$-st order have prescribed values at the boundary. Again we take $\eta(x)$ to be an arbitrary function, continuous with its derivatives up to the $2n$-th order and with $\eta(x) = \eta'(x) = \cdots = \eta^{(n-1)}(x) = 0$ at the boundary points $x = x_0$ and $x = x_1$. Just as before we obtain for the first variation $\delta J = \epsilon \frac{d}{d\epsilon} J[y + \epsilon\eta]|_{\epsilon=0}$ the expression

$$\delta J = \epsilon \int_{x_0}^{x_1} (F_y\eta + F_{y'}\eta' + \cdots + F_{y^{(n)}}\eta^{(n)})\, dx.$$

By repeated integration by parts we can eliminate all the derivatives of the function η from the integral, transforming it into the form

$$\delta J = \epsilon \int_{x_0}^{x_1} \eta \left[F_y - \frac{d}{dx} F_{y'} + \frac{d^2}{dx^2} F_{y''} - \cdots + (-1)^n \frac{d^n}{dx^n} F_{y^{(n)}} \right] dx. \tag{20}$$

Thus, by the fundamental lemma (see subsection 1), we obtain the differential equation of order $2n$,

$$[F]_y = F_y - \frac{d}{dx} F_{y'} + \frac{d^2}{dx^2} F_{y''} - \cdots + (-1)^n \frac{d^n}{dx^n} F_{y^{(n)}} = 0. \tag{21}$$

as a necessary condition for an extremum. This equation is again called Euler's equation. The $2n$ constants of integration occurring in the general solution of (21) are to be determined from the $2n$ boundary conditions.

To determine several functions y, z, $\cdots$ which solve the variational problem

$$\int_{x_0}^{x_1} F(x, y, z, \cdots, y', z', \cdots, y'', z'', \cdots)\, dx = \text{min.},$$

we have to consider the corresponding systems of Euler differential equations in the same way.

4. Several Independent Variables. The problem of determining extrema of multiple integrals leads to one or more *partial differential equations* for the required functions, just as the problems considered so far lead to ordinary differential equations. Let us consider, for example, the problem of finding an extremum of the double integral

$$J = \iint_G F(x, y, u, u_x, u_y)\, dx\, dy \tag{22}$$

over a given region of integration G by determining a suitable function u which is continuous, has continuous derivatives up to the second order, and takes on prescribed values on the boundary of G. We introduce an arbitrary function $\eta(x, y)$, on which we shall later impose the boundary condition $\eta = 0$, and obtain as a necessary condition for an extremum the vanishing of the first variation

$$\delta J = \epsilon \left(\frac{d}{d\epsilon}\, \Phi(\epsilon)\right)_{\epsilon=0} = \epsilon \left(\frac{d}{d\epsilon}\, J[u + \epsilon\eta]\right)_{\epsilon=0};$$

this is equivalent to equation

$$\delta J = \epsilon \iint_G (F_u \eta + F_{u_x} \eta_x + F_{u_y} \eta_y)\, dx\, dy = 0, \tag{23}$$

which we may again transform by integration by parts. We assume—as usual—that the boundary curve Γ of G has a tangent which turns piecewise continuously. Then according to Gauss's integral theorem[1] we have

$$\iint_G (\eta_x F_{u_x} + \eta_y F_{u_y})\, dx\, dy = \int_\Gamma \eta(F_{u_x}\, dy - F_{u_y}\, dx) - \iint_G \eta \left(\frac{\partial}{\partial x} F_{u_x} + \frac{\partial}{\partial y} F_{u_y}\right) dx\, dy.$$

We thus obtain

$$\delta J = \epsilon \iint_G \eta \left\{F_u - \frac{\partial}{\partial x} F_{u_x} - \frac{\partial}{\partial y} F_{u_y}\right\} dx\, dy + \epsilon \int_\Gamma \eta(F_{u_x}\, dy - F_{u_y}\, dx)$$

$$= \iint_G \delta u[F]_u\, dx\, dy + \int_\Gamma \delta u(F_{u_x}\, dy - F_{u_y}\, dx) = 0,$$

[1] See, for example, R. Courant, Differential and Integral Calculus, Vol. II, p. 360, 2nd ed. rev., Interscience Publishers, Inc., New York, 1947.

and if, corresponding to the assumption of fixed boundary values of u, we impose the condition $\eta = 0$ on the boundary, we obtain

$$\delta J = \epsilon \iint_G \eta \left\{ F_u - \frac{\partial}{\partial x} F_{u_x} - \frac{\partial}{\partial y} F_{u_y} \right\} dx\, dy = 0. \tag{24}$$

The equation $\delta J = 0$ must be valid for any arbitrary continuously differentiable function η. The lemma stated in subsection 1 for simple integrals also holds for multiple integrals and can be proved in the same way; we therefore conclude that $u(x, y)$ must satisfy the Euler differential equation

$$-[F]_u = \frac{\partial}{\partial x} F_{u_x} + \frac{\partial}{\partial y} F_{u_y} - F_u = 0, \tag{25}$$

or more explicitly

$$F_{u_x u_x} u_{xx} + 2F_{u_x u_y} u_{xy} + F_{u_y u_y} u_{yy} + F_{u_x u} u_x + F_{u_y u} u_y + F_{u_x x} + F_{u_y y} - F_u = 0.$$

From the manifold of all solutions of this equation a particular solution must be determined by means of the given boundary condition (*boundary value problem*).

Similarly, we obtain a system of such differential equations if there are several unknown functions to be determined, and a differential equation of order $2n$,

$$[F]_u = F_u - \frac{\partial}{\partial x} F_{u_x} - \frac{\partial}{\partial y} F_{u_y} + \frac{\partial^2}{\partial x^2} F_{u_{xx}} + \frac{\partial^2}{\partial x\, \partial y} F_{u_{xy}} + \frac{\partial^2}{\partial y^2} F_{u_{yy}} + \cdots + (-1)^n \frac{\partial^n}{\partial y^n} F_{u_{yy\cdots y}} = 0, \tag{26}$$

if the function F contains the derivatives u_x, u_y, $\cdots$, $u_{yy\cdots y}$ up to the n-th order.

We may consider the example $F = \frac{1}{2}(u_x^2 + u_y^2)$ (cf. page 172). Here the Euler equation is the same as the "potential equation"

$$\Delta u = u_{xx} + u_{yy} = 0.$$

The function $F = \frac{1}{2}(\Delta u)^2 = \frac{1}{2} u_{xx}^2 + u_{xx} u_{yy} + \frac{1}{2} u_{yy}^2$ leads to the Euler equation

$$\Delta\Delta u = u_{xxxx} + 2u_{xxyy} + u_{yyyy} = 0;$$

the same Euler equation can be obtained from the integrand $(\Delta u)^2 - c(u_{xx}u_{yy} - u_{xy}^2)$ with constant c.

The problem of *minimal surfaces*, i.e. the integrand

$$F = \sqrt{1 + z_x^2 + z_y^2},$$

leads to the Euler differential equation

$$z_{xx}(1 + z_y^2) - 2z_{xy}z_xz_y + z_{yy}(1 + z_x^2) = 0.$$

5. Identical Vanishing of the Euler Differential Expression. The Euler differential expression for an integrand $F(x, y, y', \cdots)$ may vanish identically for every admissible argument function. Since admissible argument functions can be constructed for which the quantities $y, y', \quad \cdot$ take on prescribed values at an arbitrary point x, the identical vanishing of the Euler expression $[F]_y$ for all functions y is equivalent to the identical vanishing of this expression if $x, y, y', \cdots$ are regarded as independent parameters. The same statement holds if the argument function depends on several independent variables.

The simplest case is that of the integrand $F(x, y, y')$. If the expression $F_y - F_{y'x} - F_{y'y}y' - F_{y'y'}y''$ vanishes, it follows that $F_{y'y'} = 0$ and therefore that F is of the form $F = A(x, y) + y'B(x, y)$. Then the Euler differential equation becomes simply the integrability condition

$$\frac{\partial A}{\partial y} - \frac{\partial B}{\partial x} = 0,$$

and the integral

$$\int_{x_0}^{x} F\,dx = \int_{x_0}^{x} (A + By')\,dx = \int_{(x_0,y_0)}^{(x,y)} (A\,dx + B\,dy)$$

is indeed independent of the curve of integration, according to a well-known theorem of integral calculus.[1] If the upper limit x is regarded as a variable the integral becomes a function $G(x, y)$ of the upper limit; we obtain

$$F(x, y, y') = \frac{d}{dx}G(x, y).$$

[1] We shall see presently that this independence also follows directly from the identical vanishing of the Euler expression.

Thus the Euler differential expression for F vanishes identically if and only if this relation holds.

The situation is similar for the integrand $F(x, y, y', \cdots, y^{(n)})$. In this case also a necessary and sufficient condition for the identical vanishing of the Euler differential expression $[F]_y$ is that F be representable in the form

$$F = \frac{dG}{dx}$$

where $G(x, y, y', \cdots, y^{(n-1)})$ contains only derivatives of y up to the $(n-1)$-st order.

We may verify by a simple calculation that this condition is sufficient or we may see this in the following way: The integral $\int_{x_0}^{x_1} F\,dx$ is assumed to depend only on the values of the function y and its first $n-1$ derivatives at the end points; therefore it is not changed if the function is varied in the interior, as long as these boundary values are retained. Therefore the first variation and consequently the Euler expression vanish identically.

To show that the condition is necessary we consider a family of functions $y(x, \alpha)$ with the parameter α and fixed (independent of α) boundary values of $y, y', \cdots, y^{(n-1)}$. If we denote the integral with the argument function $y(x, \alpha)$ by $J(\alpha)$, the formulas for the first variation yield

$$\frac{\partial J}{\partial \alpha} = \int_{x_0}^{x_1} [F]_y \frac{\partial y}{\partial \alpha}\,dx = 0$$

since $[F]_y$ vanishes identically. Thus J does not depend on α and is therefore a function only of the coordinates x_0 and x_1 and of the values of y and its first $n-1$ derivatives at the end points. If we suppose the initial point x_0 fixed and the upper limit x_1 variable, we obtain an equation of the form

$$\int_{x_0}^{x_1} F(x, y, y', \cdots, y^{(n)})\,dx = G(x_1, y, y', \cdots, y^{(n-1)}),$$

from which the rest of our assertion follows by differentiation with respect to the upper limit.

The case of integrals with argument functions of several variables is analogous, as long as the integrand $F(x, y, u, u_x, u_y)$ contains only

derivatives of the first order. We then obtain, in the same manner as before, the following theorem: *A necessary and sufficient condition for the Euler expression* $[F]_u$ *to vanish identically in the argument function* u *is that* F *be representable in the form*

$$F = A_x + B_y$$

where A *and* B *are functions of* x, y, *and* u. An expression of this form is known as a *divergence expression.*

A divergence expression may also be characterized by the requirement that the value of the double integral $\iint_\Omega F\,dx\,dy$ *remain unchanged if the function* u *is so varied that the variation is confined to a subregion interior to* G.

According to the Gauss integral theorem we have

$$\iint_G F\,dx\,dy = \int_\Gamma (A\,dy - B\,dx),$$

where the integral on the right is a line integral extended in the positive sense around the boundary curve Γ of G.

The situation may be somewhat more complicated if the integrand F contains partial derivatives of orders higher than the first. The theorem that the Euler expression vanishes identically if and only if F can be represented in the form

$$F = A_x + B_y$$

(i.e. if F is a divergence expression) remains valid; however, it is not in general possible to choose A and B in such a way that the derivatives occurring in these functions are of lower order than those occurring in F.

The simplest example of a divergence expression of the second order is $F = u_{xx}u_{yy} - u_{xy}^2$. We have here

$$\begin{aligned} F &= (u_x u_{yy})_x - (u_x u_{xy})_y = -(u_y u_{xy})_x + (u_y u_{xx})_y \\ &= -\tfrac{1}{2}[(u_x^2)_{yy} - 2(u_x u_y)_{xy} + (u_y^2)_{xx}]. \end{aligned}$$

Another example is given by the identity

$$\frac{u_{xx}u_{yy} - u_{xy}^2}{(1 + u_x^2 + u_y^2)^{3/2}} = \frac{\partial}{\partial y}\left[\frac{u_{xx}u_y}{(u_x^2 + 1)\sqrt{1 + u_x^2 + u_y^2}}\right] - \frac{\partial}{\partial x}\left[\frac{u_{xy}u_y}{(u_x^2 + 1)\sqrt{1 + u_x^2 + u_y^2}}\right].$$

The expression

$$\frac{u_{xx}u_{yy} - u_{xy}^2}{(1 + u_x^2 + u_y^2)^{3/2}}$$

is the Gaussian curvature of the surface $z = u(x, y)$ multiplied by $\sqrt{1 + u_x^2 + u_y^2}$; its divergence character expresses the well-known fact that the integral of the curvature over a segment of a surface, i.e. the total curvature of the segment, depends only on the tangent planes of the surface along the boundary of the segment.

Our considerations lead to the following theorem: *If the difference between the integrands of two variational problems is a divergence expression, then the Euler equations and therefore the families of extremals are identical for the two variational problems.* (See footnote, page 211.)

6. Euler Equations in Homogeneous Form. In geometrical problems, where we are concerned with determining curves or surfaces by minimum conditions, it is often appropriate to refrain from arbitrarily designating coordinates as independent variables. Instead we may employ a parametric representation $x = x(t)$, $y = y(t)$ for the curve (or $x = x(u, v)$, $y = y(u, v)$, $z = z(u, v)$ for the surface) where t(or u and v) is the independent variable, and where the equations

$$\dot{x} = \dot{y} = 0$$

$$(\text{or} \quad x_u y_v - x_v y_u = y_u z_v - y_v z_u = z_u x_v - z_v x_u = 0)$$

are not satisfied simultaneously. Here differentiation with respect to t is indicated by a dot. We begin by considering the simplest variational problem, which takes the form

$$(27) \qquad J = \int_{x_0}^{x_1} F\left(x, y, \frac{dy}{dx}\right) dx = \int_{t_0}^{t_1} \mathfrak{F}(x, y, \dot{x}, \dot{y})\, dt = \min.,$$

where

$$\mathfrak{F} = \dot{x} F\left(x, y, \frac{\dot{y}}{\dot{x}}\right).$$

This function $\mathfrak{F}$ is "homogeneous" of degree 1 in the derivatives $\dot{x}$, $\dot{y}$. which means that for all k it satisfies the homogeneity relation

$$(28) \qquad \mathfrak{F}(x, y, k\dot{x}, k\dot{y}) = k\mathfrak{F}(x, y, \dot{x}, \dot{y})$$

and consequently equation

$$\dot{x}\mathfrak{F}_{\dot{x}} + \dot{y}\mathfrak{F}_{\dot{y}} = \mathfrak{F}, \tag{29}$$

which follows from (28) by differentiating with respect to k and setting $k = 1$. If, conversely, $\mathfrak{F}$ is any homogeneous[1] function of the first degree in $\dot{x}$ and $\dot{y}$, i.e. if $\mathfrak{F}$ satisfies equation (28), then the variational problem $\int \mathfrak{F}\, dt = \text{min.}$ determines a curve independent of the choice of parameter. For, if we perform the parametric transformation $t = t(\tau)$ with $dt/d\tau > 0$, the interval $t_0 \leq t \leq t_1$ goes over into $\tau_0 \leq \tau \leq \tau_1$ and we obtain, making use of (28),

$$\begin{aligned}\int_{\tau_0}^{\tau_1} \mathfrak{F}\left(x, y, \frac{dx}{d\tau}, \frac{dy}{d\tau}\right) d\tau &= \int_{\tau_0}^{\tau_1} \mathfrak{F}\left(x, y, \dot{x}\,\frac{dt}{d\tau}, \dot{y}\,\frac{dt}{d\tau}\right) d\tau \\ &= \int_{\tau_0}^{\tau_1} \mathfrak{F}(x, y, \dot{x}, \dot{y})\,\frac{dt}{d\tau}\, d\tau \\ &= \int_{t_0}^{t_1} \mathfrak{F}(x, y, \dot{x}, \dot{y})\, dt.\end{aligned}$$

Thus the variational problem is invariant with respect to a parametric transformation that does not alter the sense of traversal; the extremal curves do not depend on the choice of parameter.

The homogeneous problem leads to the two Euler equations

$$\mathfrak{F}_x - \dot{\mathfrak{F}}_{\dot{x}} = 0, \qquad \mathfrak{F}_y - \dot{\mathfrak{F}}_{\dot{y}} = 0, \tag{30}$$

which, together with relation (29), must essentially be equivalent to the original differential equation (14), and therefore cannot be independent. We find the interdependence by deriving the following identities from (29) by differentiation:

$$\mathfrak{F}_x = \dot{x}\mathfrak{F}_{x\dot{x}} + \dot{y}\mathfrak{F}_{x\dot{y}}, \qquad \mathfrak{F}_y = \dot{x}\mathfrak{F}_{y\dot{x}} + \dot{y}\mathfrak{F}_{y\dot{y}};$$

$$\dot{x}\mathfrak{F}_{\dot{x}\dot{x}} + \dot{y}\mathfrak{F}_{\dot{x}\dot{y}} = 0, \qquad \dot{x}\mathfrak{F}_{\dot{x}\dot{y}} + \dot{y}\mathfrak{F}_{\dot{y}\dot{y}} = 0;$$

$$\mathfrak{F}_{\dot{x}\dot{x}} : \mathfrak{F}_{\dot{x}\dot{y}} : \mathfrak{F}_{\dot{y}\dot{y}} = \dot{y}^2 : -\dot{x}\dot{y} : \dot{x}^2.$$

[1] In geometrical examples the function is often only positively homogeneous (i.e. (28) holds for $k > 0$) not homogeneous in the complete sense. For, in this case traversal of the curve in the opposite direction would lead to the opposite sign for the integral, while e.g., the arc length, which is defined by $\int \sqrt{\dot{x}^2 + \dot{y}^2}\, dt$ with the positive square root, always has the same value irrespective of the direction of traversal. The above considerations also hold for such positively homogeneous integrands.

The value

$$\frac{\mathfrak{F}_{\dot{x}\dot{x}}}{\dot{y}^2} = -\frac{\mathfrak{F}_{\dot{x}\dot{y}}}{\dot{x}\dot{y}} = \frac{\mathfrak{F}_{\dot{y}\dot{y}}}{\dot{x}^2}$$

is customarily denoted by $\mathfrak{F}_1$.

From the above identities it follows that

$$\begin{aligned}\mathfrak{F}_x - \dot{\mathfrak{F}}_{\dot{x}} &= \dot{x}\mathfrak{F}_{x\dot{x}} + \dot{y}\mathfrak{F}_{x\dot{y}} - \mathfrak{F}_{\dot{x}x}\dot{x} - \mathfrak{F}_{\dot{x}y}\dot{y} - \mathfrak{F}_{\dot{x}\dot{x}}\ddot{x} - \mathfrak{F}_{\dot{x}\dot{y}}\ddot{y} \\ &= \dot{y}[\mathfrak{F}_{x\dot{y}} - \mathfrak{F}_{\dot{x}y} + (\dot{x}\ddot{y} - \dot{y}\ddot{x})\mathfrak{F}_1], \\ \mathfrak{F}_y - \dot{\mathfrak{F}}_{\dot{y}} &= -\dot{x}[\mathfrak{F}_{x\dot{y}} - \mathfrak{F}_{\dot{x}y} + (\dot{x}\ddot{y} - \dot{y}\ddot{x})\mathfrak{F}_1],\end{aligned}$$

so that the two equations (30) are connected by the identity

$$\dot{x}(\mathfrak{F}_x - \dot{\mathfrak{F}}_{\dot{x}}) + \dot{y}(\mathfrak{F}_y - \dot{\mathfrak{F}}_{\dot{y}}) = 0 \tag{31}$$

and may be replaced, for example, by the single equation

$$\mathfrak{F}_{x\dot{y}} - \mathfrak{F}_{\dot{x}y} + (\dot{x}\ddot{y} - \dot{y}\ddot{x})\mathfrak{F}_1 = 0. \tag{32}$$

The situation is quite analogous when we are concerned with the determination of several functions of one variable. Here the variational problem $\int_{x_0}^{x_1} F(x, y, z, y', z')\,dx = \text{min.}$ goes over into the problem $\int_{t_0}^{t_1} \mathfrak{F}(x, y, z, \dot{x}, \dot{y}, \dot{z})\,dt = \text{min.}$ with $\mathfrak{F} = \dot{x}F(x, y, z, \dot{y}/\dot{x}, \dot{z}/\dot{x})$; the function $\mathfrak{F}$ is homogeneous of the first degree in the variables $\dot{x}$, $\dot{y}$, and $\dot{z}$.

The advantage of the homogeneous representation is not solely that of symmetry. Thus, for example, curves on which x does not increase monotonically, such as closed curves, cannot be represented in the form $y = y(x)$; thus they cannot be readily treated by the non-homogeneous representation.

In the case of problems in several dimensions the homogeneous representation is obtained as follows: If in the integral

$$\iint F(x, y, z, z_x, z_y)\,dx\,dy$$

the variables x and y and the function z are written as functions of two parameters u and v in such a way that the Jacobian determinant of x and y with respect to u and v

$$\frac{\partial(x, y)}{\partial(u, v)} = x_u y_v - x_v y_u$$

is different from zero, then we have

$$z_x = \frac{z_u y_v - z_v y_u}{x_u y_v - x_v y_u}, \qquad z_y = -\frac{z_u x_v - z_v x_u}{x_u y_v - x_v y_u},$$

and the integral takes the form

$$\iint F(x, y, z, z_x, z_y)\, dx\, dy$$

$$(33) \qquad = \iint F\left(x, y, z, -\frac{\dfrac{\partial(y, z)}{\partial(u, v)}}{\dfrac{\partial(x, y)}{\partial(u, v)}}, -\frac{\dfrac{\partial(z, x)}{\partial(u, v)}}{\dfrac{\partial(x, y)}{\partial(u, v)}}\right) \frac{\partial(x, y)}{\partial(u, v)}\, du\, dv$$

$$= \iint \mathfrak{F}\left(x, y, z, \frac{\partial(y, z)}{\partial(u, v)}, \frac{\partial(z, x)}{\partial(u, v)}, \frac{\partial(x, y)}{\partial(u, v)}\right) du\, dv.$$

Here the integrand $\mathfrak{F}$ is homogeneous of the first degree in the last three Jacobians. The relations derived above for the case of one-dimensional integrals, in particular the identity (31) and the symmetric form (32) of the differential equation, may easily be generalized to the case of several variables. Since these generalizations will not be needed in the present book, the reader is referred to the literature.[1]

7. Relaxing of Conditions. Theorems of du Bois-Reymond and Haar. So far we have required that the comparison functions possess continuous derivatives up to the highest order occurring in the Euler differential equation. This requirement appears unnaturally restrictive from the point of view of the variational problem; for example, the variational problem with the integrand $F(x, y, y')$ has a meaning even if the first derivative is required to be only piecewise continuous and no assumptions at all regarding the second derivative are made. It is a priori conceivable that if the conditions of admissibility are broadened in this way one might obtain a new solution which no longer satisfies the Euler differential equation.

[1] O. Bolza, Vorlesungen über Variationsrechnung, pp. 666-671, B. G. Teubner, Leipzig and Berlin, 1909; G. Kobb, Sur les maxima et minima des intégrales doubles, Acta Math., Vol. 16, 1892-3, pp. 65-140.

Let us first consider an actual minimum problem and assume that $y(x)$ is that function with continuous first and second derivatives which yields the minimum. *Then the function $y(x)$ yields the minimum even if we enlarge the space of admissible functions to include functions y^* which need not have second derivatives.* For, according to the Weierstrass approximation theorem, we can approximate the derivative $y^{*\prime}$ by a polynomial $p'(x)$ and the function y^* by a polynomial $p(x)$ as precisely as desired, where $p(x)$ satisfies the boundary conditions $p(x_0) = y_0$, $p(x_1) = y_1$.[1] Then $J[p]$ also differs arbitrarily little from $J[y^*]$. But since $p(x)$ is an admissible comparison function with continuous second derivatives we have $J[p] \geq J[y]$, and therefore $J[y^*] \geq J[y]$.

If the minimum problem, or more generally the problem of stationary values, is formulated with these broadened admissibility conditions, we stipulate only that the solution must possess a continuous first derivative. We now pose the question: does this function automatically possess any higher order derivatives? If so, does it satisfy the Euler differential equation? This question is answered in the affirmative by the following theorem of du Bois-Reymond: *Given a variational problem with integrand $F(x, y, y')$, suppose F has continuous first and second derivatives with respect to all its arguments. Let $y(x)$ be a continuous function with prescribed values at the boundary and with a continuous first derivative. Assume that the first variation* (15) *vanishes for $y(x)$ and for every function $\eta(x)$ with vanishing boundary values and a continuous first derivative and that moreover $F_{y'y'} \neq 0$, then $y(x)$ has continuous second derivatives and satisfies Euler's differential equation; i.e. the vanishing of the first variation implies that $y(x)$ possesses a continuous second derivative and satisfies the Euler equation.*

[1] According to the Weierstrass theorem we may construct a polynomial $q'(x)$ which differs by less than $(\epsilon/2)(x_1 - x_0)$ from the function $y^{*\prime}(x)$ for $x_0 \leq x \leq x_1$, where ϵ is an arbitrarily fixed positive quantity. Then the polynomial

$$q(x) = y_0 + \int_{x_0}^{x_1} q'(t)\, dt$$

assumes the prescribed initial value y_0 and differs from $y^*(x)$ by at most $\epsilon/2$ in the interval. To obtain a polynomial which also takes on the required final value y_1 we simply add to $q(x)$ the linear function $l(x) = [y_1 - q(x_1)] \cdot (x - x_0)/(x_1 - x_0)$ and verify immediately that $p(x) = q(x) + l(x)$ has all the properties specified in the text.

We preface the proof by a simple lemma: *If $\varphi(x)$ is a function piecewise continuous*[1] *in the interval of integration, and if the equation*

$$\int_{x_0}^{x_1} \varphi(x)\eta(x)\,dx = 0$$

holds for arbitrary continuous functions $\eta(x)$ satisfying the condition

$$\int_{x_0}^{x_1} \eta(x)\,dx = 0$$

then $\varphi(x)$ is a constant. To prove this we first note that the above relation is certainly satisfied for constant φ. We now determine a constant c such that $\int_{x_0}^{x_1} (\varphi - c)\,dx = 0$ for our given φ; we then have $\int_{x_0}^{x_1} (\varphi - c)\eta\,dx = 0$ because $\int_{x_0}^{x_1} \varphi\eta\,dx = 0$. But now we may set $\eta = \varphi - c$, obtaining the equation $\int_{x_0}^{x_1} (\varphi - c)^2\,dx = 0$, from which our assertion follows immediately.

The following more general theorem is proved in exactly the same way: *If $\varphi(x)$ is a piecewise continuous function which satisfies the condition $\int_{x_0}^{x_1} \varphi\eta\,dx = 0$ for all continuous functions $\eta(x)$ satisfying the conditions*

$$\int_{x_0}^{x_1} \eta\,dx = 0, \qquad \int_{x_0}^{x_1} x\eta\,dx = 0, \qquad \cdots, \qquad \int_{x_0}^{x_1} x^n\eta\,dx = 0$$

then φ is a polynomial of the n-th degree:

$$\varphi = c_0 + c_1x + \cdots + c_nx^n.$$

To prove du Bois-Reymond's theorem we note that the equation

$$\int_{x_0}^{x_1} (F_v\zeta + F_{v'}\zeta')\,dx = 0$$

holds for any continuously differentiable function $\zeta(x)$ for which

$$\zeta(x_0) = \zeta(x_1) = 0.$$

[1] See Ch. II, p. 48.

We introduce the abbreviations $F_y = A'$, $F_{y'} = B$, $\int_{x_0}^{x} F_y\,dx = A$ and obtain, integrating by parts,

$$\int_{x_0}^{x_1} (A'\zeta + B\zeta')\,dx = \int_{x_0}^{x_1} \zeta'(B - A)\,dx = 0.$$

We choose an arbitrary function $\zeta' = \eta$, subject only to the conditions that it must be continuous, that $\int_{x_0}^{x_1} \eta\,dx = \zeta(x_1) - \zeta(x_0) = 0$ holds, and that the boundary conditions on ζ must be satisfied. Applying the preceding simple lemma we obtain

$$B - A = F_{y'} - \int_{x_0}^{x} F_y\,dx = c, \tag{34}$$

where c is independent of x. This equation takes the place of Euler's equation. But since $\int_{x_0}^{x} F_y dx$ is differentiable with respect to the upper limit x and since c also is, it follows that $F_{y'}$ is differentiable; therefore Euler's equation

$$\frac{d}{dx} F_{y'} - F_y = 0 \tag{34a}$$

holds. Now if F is twice continuously differentiable with respect to its arguments, and if furthermore the Legendre condition $F_{y'y'} \neq 0$ is fulfilled, it follows that the piecewise continuous function y' is actually continuous and possesses a continuous derivative. For, in the first place, since $F_{y'y'} \neq 0$, y' may be expressed as a continuously differentiable function $\varphi(x, y, F_{y'})$. Since $F_{y'}$ is a continuous function of x in virtue of (34), the same must be true of y'. Thus the arguments y and $F_{y'}$ of φ are continuously differentiable and the same is therefore true of $\varphi = y'$.

Du Bois-Reymond's result may immediately be extended to an integrand of the form $F(x, y, y', \cdots, y^{(n)})$ by using the generalized version of the above lemma. The details may be left to the reader.

In the case of variational problems with several independent variables the situation is a little more subtle. If in the problem with integrand $F(x, y, u, u_x, u_y)$ the space of admissible functions is extended to include all continuous functions with piecewise continuous derivatives, it is no longer true that the vanishing of the first variation

necessarily implies that the second derivatives exist and are continuous and that Euler's differential equation holds. However, for several dimensions there does exist an analogue to du Bois-Reymond's theorem (theorem of Haar): *The vanishing of the first variation of the integral of* $F(x, y, u, u_x, u_y)$ *for continuous* u *with continuous derivatives* u_x *and* u_y *is equivalent to equation*

$$\iint_B F_u \, dx \, dy = \int_R (F_{u_x} \, dy - F_{u_y} \, dx), \tag{35}$$

where the integral on the left is taken over any arbitrary simply connected subregion B *of* G *bounded by piecewise smooth curves, and the integral on the right is extended over the boundary* R *of* B *in the positive sense.* In the particular case where F does not depend explicitly on u, Haar's theorem states that the integral on the right vanishes for a closed curve R; this is equivalent to the statement that a function $\Phi(x, y)$ exists for which the system of differential equations

$$F_{u_x} = \Phi_y, \qquad F_{u_y} = -\Phi_x$$

holds in every simply connected subregion of G. Therefore the above integral relation or this system of differential equations of the first order takes the place of the Euler second order differential equation.

To prove Haar's theorem we need only show that the integral relation is valid for the special case that B is a square. It is then immediately valid for any region consisting of a finite number of squares, and the theorem for an arbitrary region follows in the usual manner. Let us consider the square B: $x_0 \leq x \leq x_1$, $y_0 \leq y \leq y_1$. The vanishing of the first variation for B is expressed by the equation

$$\iint_B (F_u \zeta + F_{u_x} \zeta_x + F_{u_y} \zeta_y) \, dx \, dy = 0$$

if ζ vanishes on the boundary of the square. We now specialize the variation $\zeta(x, y)$ to the form $\zeta(x, y) = v(x)w(y)$, where $v(x)$ vanishes for $x = x_0, x_1$ and $w(y)$ vanishes for $y = y_0, y_1$. Thus we obtain the equation

$$\int_{y_0}^{y_1} \int_{x_0}^{x_1} (F_u vw + F_{u_x} v'w + F_{u_y} vw') \, dx \, dy = 0,$$

from which the desired result follows by repeated application of du Bois-Reymond's theorem: We introduce the abbreviations

$$F_{u_x} = A_y(x, y), \qquad F_{u_y} = B_x(x, y), \qquad F_u = C_{xy}(x, y),$$

$$\int_{y_0}^{y} F_{u_x}\, dy = A(x, y), \qquad \int_{x_0}^{x} F_{u_y}\, dx = B(x, y), \qquad \int_{x_0}^{x} F_u\, dx = C_y(x, y),$$

$$\int_{y_0}^{y}\int_{x_0}^{x} F_u\, dx\, dy = C(x, y)$$

and integrate by parts; we obtain

$$\int_{y_0}^{y_1} dy\left\{\int_{x_0}^{x_1} (-\, C_y v' w + A_y v' w - B v' w')\, dx\right\} = 0$$

or

$$\int_{x_0}^{x_1} dx\, v'\left\{\int_{y_0}^{y_1} (-\, C_y w + A_y w - B w')\, dy\right\} = 0.$$

Since v' is the derivative of an arbitrary function which vanishes on the boundary, it follows from the above lemma that

$$\int_{y_0}^{y_1} (-\, C_y w + A_y w - B w')\, dy = c,$$

where c is independent of x; integration by parts transforms this to

$$\int_{y_0}^{y_1} (C - A - B) w'\, dy = c.$$

In this equation we set x equal first to x_1 and then to x_0 and subtract one of the resulting equations from the other; we obtain

$$\int_{y_0}^{y_1} [D(x_1, y) - D(x_0, y)] w'\, dy = 0,$$

where $D = C - A - B$. Applying du Bois-Reymond's theorem once more we have

$$D(x_1, y_1) - D(x_0, y_1) = D(x_1, y_0) - D(x_0, y_0),$$

which is just the equation

$$\int_{y_0}^{y_1}\int_{x_0}^{x_1} F_u\, dx\, dy = \int (F_{u_x}\, dy - F_{u_y}\, dx) \tag{35'}$$

that was to be proved.

8. Variational Problems and Functional Equations. So far we have been concerned with variational problems involving functionals formed by integrating a given differential expression in the argument function. However, more general classes of functionals are often encountered in variational problems. We shall show by a few examples how the above pattern may be applied to obtain functional equations which replace the Euler differential equations.

(a) We want to make the expression

$$J[\varphi] = \iint K(s, t)\varphi(s)\varphi(t)\, ds\, dt + \int [\varphi(s)]^2\, ds - 2\int \varphi(s)f(s)\, ds$$

stationary, where $K(s, t)$ is a given continuous symmetric function of s and t, $f(s)$ a given continuous function of s, and $\varphi(s)$ the unknown continuous argument function. All integrations are to be extended over a given interval $a \leq s \leq b$, $a \leq t \leq b$. If we replace φ by $\varphi + \epsilon\zeta$ and consider $J[\varphi + \epsilon\zeta] = \Phi(\epsilon)$ as a function of ϵ, we obtain, after a simple transformation,

$$\delta J = \epsilon \frac{d\Phi}{d\epsilon}\bigg|_{\epsilon=0} = 2\epsilon \int_a^b \zeta(t)\left[\int_a^b K(s, t)\varphi(s)\, ds + \varphi(t) - f(t)\right] dt.$$

Thus the requirement $\delta J = 0$ leads to the Fredholm integral equation

$$\int_a^b K(s, t)\varphi(s)\, ds + \varphi(t) - f(t) = 0$$

as the Euler equation of the problem.

The extremum problems treated in Chapter III for integral equations with symmetric kernels $K(s, t)$ likewise can be easily treated by our variational scheme.

(b) The expression

$$J[\varphi] = \int_{-\infty}^{\infty} [p(x)(\varphi'(x))^2 + 2\varphi(x + 1)\varphi(x - 1) - \varphi^2(x) - 2\varphi(x)f(x)]\, dx$$

is to be made stationary, where the argument function is continuous and has a piecewise continuous derivative in the entire interval $-\infty < x < \infty$. Forming the first variation we obtain, after a simple transformation,

$$\delta J = \epsilon \frac{d}{d\epsilon} J[\varphi + \epsilon\zeta]\bigg|_{\epsilon=0}$$

$$= 2\epsilon \int_{-\infty}^{\infty} \zeta(x)[-(p\varphi')' + \varphi(x + 2) + \varphi(x - 2) - \varphi(x) - f(x)]\, dx;$$

the Euler functional equation which expresses the vanishing of the first variation for arbitrary ζ is

$$(p\varphi')' - \varphi(x + 2) - \varphi(x - 2) + \varphi(x) + f(x) = 0.$$

Thus we obtain a differential-difference equation rather than a differential equation.

§4. *Integration of the Euler Differential Equation*

In Volume II a systematic method of integrating the Euler differential equations will be developed by means of the Hamilton-Jacobi theory. At this point we shall consider briefly the integration of the simple preceding examples. We confine ourselves to the problem $\int_{x_0}^{x_1} F(x, y, y')\,dx = \min.$

If the function F does not involve the derivative y', the Euler equation reduces to $F_y = 0$; this equation determines the function $y(x)$ implicitly. We note that in this case the boundary values cannot be prescribed arbitrarily if the problem is to have a solution.

If the function F does not contain the dependent variable y, we obtain immediately $dF_{y'}/dx = 0$, i.e. $F_{y'} = \text{const.} = c$; therefore $y' = \varphi(x, c)$ and $y = \int \varphi(x, c)\,dx$. The Euler equation can then be solved by quadrature.

If the function F does not contain the independent variable x, we can again achieve the integration by quadrature. For then

$$(y'F_{y'} - F)' = y''F_{y'} + y'F'_{y'} - F_{y'}y'' - F_y y' = y'(F'_{y'} - F_y) = 0;$$

thus it follows immediately from the Euler equation that

$$F(y, y') - y'F_{y'}(y, y') = c,$$

from which y' may be obtained as a function $\varphi(y, c)$ of y and c, and

$$x = \int \frac{dy}{\varphi(y, c)}.$$

Alternatively, we may obtain this result, at least formally, by reduction to the previous case, noting that the extremal curve leads to the vanishing of the first variation if y is regarded as the independent and x as the dependent variable. If differentiation with respect to y

is denoted by a dot we obtain the variational problem of extremizing $\int F(y, 1/\dot{x})\dot{x}\,dy$, in which the new dependent variable does not occur.

Examples (b), (c), and (d) on page 187 can now be integrated:

(b) with $\psi = 1/\sqrt{y}$, that is, $F = \sqrt{\dfrac{1 + y'^2}{y}}$;

$$y'F_{y'} - F = \frac{-1}{\sqrt{y(1 + y'^2)}} = \text{const.} = \frac{1}{c}.$$

Setting $y = \frac{1}{2}c^2(1 - \cos t)$, we have $y' = \sqrt{\dfrac{c^2 - y}{y}} = \cot(t/2)$

$$x = \int \frac{dy}{y'} = \int \tan(t/2)\,\frac{dy}{dt}\,dt$$

$$= c^2 \int \sin^2(t/2)\,dt = c_1 + \tfrac{1}{2}c^2(t - \sin t).$$

Thus the brachistochrones are the cycloids described by a point on the circumference of a circle of radius $c^2/2$ which rolls on the x-axis.

(c) $F = y\sqrt{1 + y'^2}$; $\quad y'F_{y'} - F = \dfrac{-y}{\sqrt{1 + y'^2}} = -\dfrac{1}{c}$,

$$y = \frac{1}{c}\cosh(cx + c_1).$$

The surface of revolution of least area connecting two given circles is thus obtained by revolving a catenary about its axis.

(d) $F = y\sqrt{1 - \dot{y}^2}$; $\quad \dot{y}F_{\dot{y}} - F = \dfrac{-y}{\sqrt{1 - \dot{y}^2}} = -\dfrac{1}{c}$,

$$y = \frac{1}{c}\sin(cs + c_1).$$

The other coordinate x is given by

$$x = \int \sqrt{1 - \dot{y}^2}\,ds = \int \sin(cs + c_1)\,ds = -\frac{1}{c}\cos(cs + c_1) + c_2;$$

thus the solution of the isoperimetric problem can only be a circle.

§5. *Boundary Conditions*

In the previous sections we have postulated that the functions to be determined assume prescribed values at the boundary of the region of integration. However, in many problems there are no a priori conditions for the boundary values, or the behavior of the function at the boundary may be restricted by more general conditions. *If no boundary conditions are prescribed for the unknown functions in a fixed basic region we speak of free boundary values.* Problems also arise in which the boundaries themselves are free subject to suitable restrictions. In geometry, for example, we often wish to determine curves with end points on a prescribed curve or surface, or surfaces with boundaries on a given surface. Here it is part of the problem to determine the region of integration of the independent variables. Problems of these types may be treated by a simple generalization of our previous procedure: the expression for the first variation δJ of the integral J is adapted to our more general problems since the variations of the functions at the boundary do not necessarily have to be assumed equal to zero.

1. Natural Boundary Conditions for Free Boundaries. We consider the variational problem of §3, 1 with the integral

$$J = \int_{x_0}^{x_1} F(x, y, y')\, dx$$

but no longer impose conditions on the argument function $y(x)$ at $x = x_0, x_1$. The necessary condition for J to be stationary is that the first variation

$$\delta J = F_{y'}\delta y \Big|_{x_0}^{x_1} + \int_{x_0}^{x_1} [F]_y \delta y\, dx$$

(equation (15), page 186) vanish. It is evident that the Euler equation $[F]_y = 0$ must be satisfied. For, if J is stationary with respect to variations which do not have prescribed boundary values, then it is certainly stationary with respect to the smaller class of variations for which $\delta y = 0$ on the boundary which implies Euler's equation. Therefore we need only consider that part of δy which depends on the boundary. Because of the arbitrariness of δy at the boundary we obtain as a necessary condition the "natural boundary condition"

$$F_{y'} = 0 \qquad \text{for } x = x_0 \text{ and } x = x_1 .$$

In the same way, from the expressions for the first variation (pages 188–191) we obtain as necessary conditions for the stationary character of the integrals

$$\int_{x_0}^{x_1} F(x, y, z, \cdots, y', z', \cdots)\, dx, \tag{36}$$

$$\iint_G F(x, y, u, u_x, u_y)\, dx\, dy, \tag{37}$$

$$\iint_{\square} F(x, y, u, u_x, u_y, v, v_x, v_y, \cdots)\, dx\, dy, \tag{38}$$

respectively, the natural boundary conditions

$$F_{y'} = F_{z'} = 0 \qquad \text{for } x = x_0 \text{ and } x = x_1,$$

$$F_{u_x}\frac{dy}{ds} - F_{u_y}\frac{dx}{ds} = 0,$$

$$F_{u_x}\frac{dy}{ds} - F_{u_y}\frac{dx}{ds} = 0, \qquad F_{v_x}\frac{dy}{ds} - F_{v_y}\frac{dx}{ds} = 0, \cdots$$

in addition to the Euler equations. The last two sets of conditions are to be satisfied on the boundary Γ of G whose arc length is denoted by s.

The concept of natural boundary conditions is significant because it is easily applied to more general types of variational problems, including those in which boundary values occur explicitly. The following important examples illustrate this:

$$J = \int_{x_0}^{x_1} F(x, y, y')\, dx - \varphi(y_0) + \psi(y_1), \tag{39}$$

$$y_0 = y(x_0),\ y_1 = y(x_1) \qquad \text{(not prescribed)}$$

and

$$J = \iint F(x, y, u, u_x, u_y)\, dx\, dy + \int_{\Gamma} \Phi(s, u, u_s)\, ds \qquad \left(u_s = \frac{du}{ds}\right). \tag{40}$$

The variations are given by

$$\delta J = \int_{x_0}^{x_1} [F]_y\, \delta y\, dx + [\psi'(y_1) + F_{y'}(x_1, y(x_1), y'(x_1))]\, \delta y_1 - [\varphi'(y_0) + F_{y'}(x_0, y(x_0), y'(x_0))]\, \delta y_0 \tag{41}$$

and

$$\delta J = \iint_G [F]_u \delta u \, dx \, dy + \int_\Gamma \left(F_{u_x} \frac{dy}{ds} - F_{u_y} \frac{dx}{ds} + [\Phi]_u \right) \delta u \, ds, \tag{42}$$

respectively, with

$$[\Phi]_u = \Phi_u - \frac{d}{ds} \Phi_{u_s}. \tag{43}$$

The corresponding natural boundary conditions are:

$$[F_{y'} + \varphi'(y)]|_{x_0} = 0, \qquad [F_{y'} + \psi'(y)]|_{x_1} = 0;^1$$

$$F_{u_x} \frac{dy}{ds} - F_{u_y} \frac{dx}{ds} + \Phi_u - \frac{d}{ds} \Phi_{u_s} = 0.$$

In the particular case

$$J = \iint_G (u_x^2 + u_y^2) \, dx \, dy + \int_\Gamma \sigma u^2 \, ds \tag{44}$$

with a continuous boundary function $\sigma(s)$ the expression for the variation is

$$\delta J = -2 \iint_G (u_{xx} + u_{yy}) \, \delta u \, dx \, dy + 2 \int_\Gamma \left(\frac{\partial u}{\partial n} + \sigma u \right) \delta u \, ds, \tag{45}$$

where $\partial/\partial n$ denotes differentiation in the direction of the outward normal. For the more general integral

$$J = \iint_G [p(u_x^2 + u_y^2) - q u^2] \, dx \, dy + \int_\Gamma p \sigma u^2 \, ds, \tag{44'}$$

in which $p(x, y)$ and its first derivatives are continuous in G, $q(x, y)$ is continuous in G, and $\sigma(s)$ is continuous on Γ, we obtain similarly

$$\delta J = -2 \iint_G [(p u_x)_x + (p u_y)_y + q u] \, \delta u \, dx \, dy + 2 \int_\Gamma p \left(\frac{\partial u}{\partial n} + \sigma u \right) \delta u \, ds. \tag{45'}$$

If in (39) and (41) we set

$$\varphi(y) = l(y - a)^2, \qquad \psi(y) = (y - b)^2,$$

[1] The notation $|_{x_0}$ means that the expression at the left is to be evaluated at $x = x_0$.

the natural boundary conditions become

$$\frac{1}{2l} F_{y'}\Big|_{x_0} + y_0 - a = 0, \qquad \frac{1}{2l} F_{y'}\Big|_{x_1} + y_1 - b = 0.$$

The passage to the limit $l \to \infty$ yields the condition for fixed boundary values

$$y_0 = a, \qquad y_1 = b,$$

so that *the simplest variational problem with fixed end points of the extremals appears as a limiting case of a problem with free boundaries.*

In general we can, by adding boundary terms or boundary integrals,[1] *essentially modify the natural boundary conditions without altering the Euler equations.*

2. Geometrical Problems. Transversality. In problems where the end points of the required curve are free on given curves or surfaces,[2] or, more generally, where the boundary of the region of integration is not fixed, parametric representation is useful. We shall derive the boundary conditions for an extremum if the unknown plane curve $y(x)$ begins on a fixed curve $T(x, y) = 0$, while the end point at x_1 is fixed. We introduce a parameter t varying between the fixed limits $t_0 \leq t \leq t_1$ and transform $J = \int_{x_0}^{x_1} F(x, y, y')\, dx$ into the integral $J = \int_{t_0}^{t_1} \mathfrak{F}(x, y, \dot{x}, \dot{y})\, dt$, with $\mathfrak{F} = \dot{x}F(x, y, \dot{y}/\dot{x})$. The initial condition is $T(x(t_0), y(t_0)) = 0$; the values $x(t_1)$ and $y(t_1)$ are fixed. We have thus eliminated the bothersome variability of the interval of integration. We now introduce two functions $\xi(t)$, $\eta(t)$ which vanish at $t = t_1$ but are otherwise arbitrary, and two paramcters ϵ_1, ϵ_2 satisfying the condition

$$\Psi(\epsilon_1, \epsilon_2) = T[x(t_0) + \epsilon_1\xi(t_0), y(t_0) + \epsilon_2\eta(t_0)] = 0.$$

[1] Instead of adding such boundary integrals one might add divergence expressions to the integrand over the entire region of integration (see pages 195 and 251, footnote).

[2] The "free boundaries" just considered are, of course, a special case of such problems. Thus, for example, the problem

$$\int_{x_0}^{x_1} F(x, y, y')\, dx = \text{min.},$$

where the values $y(x_0)$, $y(x_1)$ may be arbitrary, may also be formulated as follows: Find the curve with end points on the vertical straight lines $x = x_0$, $x = x_1$ which minimizes the integral.

If our curve is extremal, the function

$$\Phi(\epsilon_1, \epsilon_2) = \int_{t_0}^{t_1} \mathfrak{F}(x + \epsilon_1\xi, y + \epsilon_2\eta, \dot{x} + \epsilon_1\dot{\xi}, \dot{y} + \epsilon_2\dot{\eta})\, dt$$

is stationary for $\epsilon_1 = 0$, $\epsilon_2 = 0$ when ϵ_1 and ϵ_2 are subjected to the condition $\Psi(\epsilon_1, \epsilon_2) = 0$. According to the theory of ordinary extrema there exist two constants λ_0, λ, not both equal to zero, such that

$$\frac{\partial}{\partial\epsilon_1}(\lambda\Psi + \lambda_0\Phi)\Big|_{\epsilon_1=\epsilon_2=0} = 0, \qquad \frac{\partial}{\partial\epsilon_2}(\lambda\Psi + \lambda_0\Phi)\Big|_{\epsilon_1=\epsilon_2=0} = 0.$$

We assume that $\partial T/\partial x$ and $\partial T/\partial y$ do not both vanish at $t = t_0$. Then we may take $\lambda_0 = 1$. Since the functions $x(t)$, $y(t)$ must satisfy the Euler equations, we find from our expressions for the first variations that the equations

$$\xi(\lambda T_x - \mathfrak{F}_{\dot{x}}) = 0, \qquad \eta(\lambda T_y - \mathfrak{F}_{\dot{y}}) = 0$$

hold at $t = t_0$. Eliminating λ we obtain the *transversality condition*

$$\mathfrak{F}_{\dot{x}}T_y - \mathfrak{F}_{\dot{y}}T_x = 0. \tag{46}$$

If the end point is also variable on a given curve a corresponding condition must, of course, be satisfied there.

The transversality condition is a relation between the direction of the extremal curve and that of the given boundary curve. It is linear in T_x and T_y; therefore, if the direction of the extremal is given, that of the boundary curve is certainly determined uniquely. (The converse is not necessarily true.) *For every given boundary curve it is possible to construct a one-parameter family of transverse extremals by drawing through each point of the boundary curve a curve which is a solution of the Euler equation and starts in the transverse direction.*

Returning to the inhomogeneous representation $y = f(x)$ of the curve we obtain the transversality condition in the form

$$(F - y'F_{y'})T_y - F_{y'}T_x = 0 \tag{47}$$

since

$$\mathfrak{F}_{\dot{x}} = F - \frac{\dot{y}}{\dot{x}}F_{y'} = F - y'F_{y'}; \qquad \mathfrak{F}_{\dot{y}} = F_{y'}; \tag{48}$$

if the boundary curve is given in the form $y = g(x)$, the transversality condition becomes

$$F + (g' - y')F_{y'} = 0.$$

Note that the latter formulation fails whenever the tangent of the boundary curve is parallel to the y-axis at the point under consideration. In this case we see from (47) that we return to the natural boundary condition $F_{y'} = 0$.

The situation is very similar in the problem of determining a space curve $y = y(x), z = z(x)$ which begins on a given surface $T(x, y, z) = 0$, passes through a given point (x_1, y_1, z_1), and renders the integral $J = \int_{x_0}^{x_1} F(x, y, z, y', z')\,dx$ stationary. Just as before we introduce a parametric representation with $\mathfrak{F}(x, y, z, \dot{x}, \dot{y}, \dot{z}) = \dot{x}F(x, y, z, \dot{y}/\dot{x}, \dot{z}/\dot{x})$ and obtain the transversality conditions

$$T_x : T_y : Tz = \mathfrak{F}_{\dot{x}} : \mathfrak{F}_{\dot{y}} : \mathfrak{F}_{\dot{z}} \tag{49}$$

or, in inhomogeneous notation, the conditions

$$T_x : T_y : T_z = (F - y'F_{y'} - z'F_{z'}) : F_{y'} : F_{z'}\,. \tag{50}$$

Once again these conditions assign one (or more) directions to each point of the boundary surface $T = 0$; therefore for each boundary surface we have a two-parameter family of extremals. To every direction of the extremal there corresponds exactly one transverse direction of the surface.

It goes without saying that the same transversality conditions apply to the end point of the curve if this point is variable on a surface.

In the case of the geodesic curves on a surface or of the shortest curves in space, transversality coincides with orthogonality. Thus for $F = \sqrt{1 + y'^2 + z'^2}$ the transversality condition is $T_x : T_y : T_z = 1 : y' : z'$. For $F = \sqrt{e + 2fy' + gy'^2}$ we obtain

$$T_x : T_y = (e + fy') : (f + gy'),$$

which is the condition for an orthogonal intersection.

Therefore, if on a surface we draw the pencil of geodesics from a point P, this pencil intersects its orthogonal trajectories transversely. Let Q be any point near P. As Q moves along an orthogonal trajectory, the length of the geodesic from P to Q is stationary. Hence this length is

constant and the trajectories, the so-called geodesic circles, are closed curves.

In the second volume we shall treat in more detail the connection between transversals and extremals. Here we merely point out that in the case of the propagation of light the *transversals* are simply the *wave fronts* of the light waves and the *extremals* are the *light rays*. By a transversal we mean a curve or surface which is everywhere transverse to a family of extremals.

§6. *The Second Variation and the Legendre Condition*

The Euler differential equation is a necessary condition for an extremum. It turns out that a particular extremal satisfying given boundary conditions can furnish an actual extremum only if it satisfies certain additional necessary conditions, which take the form of inequalities. The formulation of such inequalities, together with their refinement into sufficient conditions, is an important part of the classical calculus of variations. This will be discussed in the second volume; at present we merely state Legendre's necessary criterion:

If the extremal $\varphi = u(x)$ makes the integral $J[\varphi] = \int_{x_0}^{x_1} F(x, \varphi, \varphi')\, dx$ a minimum with respect to continuous comparison functions $\varphi(x)$ with piecewise continuous first derivatives, then the condition

$$F_{\varphi'\varphi'}(x, u, u') \geq 0$$

is satisfied everywhere along the extremal.

To prove this we expand the expression

$$J[\varphi] = \int_{x_0}^{x_1} F(x, \varphi, \varphi')\, dx$$

by Taylor's theorem:

$$J[\varphi + \epsilon\eta] = J[\varphi] + \epsilon\, J_1[\varphi, \eta] + \tfrac{1}{2}\epsilon^2\, J_2[\bar{\varphi}, \eta].$$

Here

$$J_1[\varphi, \eta] = \int_{x_0}^{x_1} (F_\varphi \eta + F_{\varphi'}\eta')\, dx,$$

$$J_2[\bar{\varphi}, \eta] = \int_{x_0}^{x_1} (\bar{F}_{\varphi\varphi}\eta^2 + 2\bar{F}_{\varphi\varphi'}\eta\eta' + \bar{F}_{\varphi'\varphi'}\eta'^2)\, dx,$$

where the bar means that in the expressions $F_{\varphi\varphi}$, $F_{\varphi\varphi'}$, $F_{\varphi'\varphi'}$ the arguments φ, φ' are to be replaced by $\bar{\varphi} = \varphi + \rho\eta$, $\bar{\varphi}' = \varphi' + \rho\eta'$, ρ being a number between 0 and ϵ. Since J is stationary for $\varphi = u$, $J_1[u, \eta]$ vanishes and a necessary condition for a minimum is evidently $J_2[\bar{\varphi}, \eta] \geq 0$ for arbitrarily chosen η.

If in $J_2[\bar{\varphi}, \eta]$ we let the parameter ϵ tend to zero, J_2 goes over into the integral

$$J_2[\varphi, \eta] = \int_{x_0}^{x_1} (F_{\varphi\varphi}\eta^2 + 2F_{\varphi\varphi'}\eta\eta' + F_{\varphi'\varphi'}\eta'^2)\, dx$$

and we obtain the necessary condition for the extremal u

$$J_2[u, \eta] \geq 0.$$

If we define the "*second variation*" $\delta^2 J$ of J by

$$\delta^2 J = \frac{\epsilon^2}{2} J_2[u, \eta]$$

the condition becomes

$$\delta^2 J \geq 0.$$

From this integral condition we obtain the above differential condition of Legendre by making use of the arbitrariness of η. We choose for η a special piecewise linear function which is different from zero only in the vicinity of the point $x = \alpha$, i.e. we choose

$$\eta = \sqrt{\sigma}\left(1 + \frac{x - \alpha}{\sigma}\right) \quad \text{for} \quad \alpha - \sigma \leq x \leq \alpha,$$

$$\eta = \sqrt{\sigma}\left(1 - \frac{x - \alpha}{\sigma}\right) \quad \text{for} \quad \alpha \leq x \leq \alpha + \sigma,$$

$$\eta = 0 \qquad \text{everywhere else.}$$

The integral $J_2[u, \eta]$ then reduces to an integral over the interval $\alpha - \sigma \leq x \leq \alpha + \sigma$, and in this interval $\eta'^2 = 1/\sigma$. If we now let σ tend to zero the first two terms of the integral tend to zero, while the limit of the third term is the value of $2F_{\varphi'\varphi'}$ at $x = \alpha$. Thus this value must be non-negative, and the necessity of the Legendre condition is established.

In the case of several unknown functions $\varphi, \psi, \cdots$ the correspond-

ing Legendre condition is: the quadratic form whose coefficient matrix is

$$\begin{pmatrix} F_{\varphi'\varphi'} & F_{\varphi'\psi'} & \cdots \\ F_{\psi'\varphi'} & F_{\psi'\psi'} & \cdots \\ \cdots & \cdots & \cdots \end{pmatrix}$$

must not assume negative values; i.e., it must be positive definite.

Instead of this form of the Legendre condition with the sign $\geq$ the more stringent Legendre condition

$$F_{\varphi'\varphi'} > 0$$

is often considered. If this condition is not only fulfilled for the extremal $\varphi = u$, but also holds for arbitrary values of x and u in a given region and for completely arbitrary u', we speak of the *strong Legendre condition*.

If in addition to this condition the more restrictive inequality

$$F_{\varphi'\varphi'}F_{\varphi\varphi} - F^2_{\varphi\varphi'} \geq 0$$

holds for all φ and x in a given region and for arbitrary φ', then the quadratic form in the integrand of J_2 is positive definite, and therefore an extremal in the given region certainly furnishes a minimum. This simple but very rough *sufficient criterion* will be refined in Volume II.

§7. *Variational Problems with Subsidiary Conditions*.

In the problems studied so far the argument functions could be chosen arbitrarily, subject only to boundary conditions, and the solution of the variational problem was determined from the Euler equations with given or natural boundary conditions. We shall now consider problems in which additional conditions are imposed on the argument functions; these conditions refer to the entire course of the argument functions and lead to essential modifications of the Euler differential equations themselves.

1. Isoperimetric Problems. A simple example is the generalized isoperimetric problem (see §1, 3(d)): Find the function y which makes the integral

$$J = \int_{x_0}^{x_1} F(x, y, y')\, dx$$

stationary, has given boundary values $y(x_0) = y_0$, $y(x_1) = y_1$, and is subject to the subsidiary condition

$$K = \int_{x_0}^{x_1} G(x, y, y')\, dx = c \tag{51}$$

with constant c.

Let us suppose that $y = y(x)$ is the desired extremal. We consider the family of neighboring curves $y + \delta y = y(x) + \epsilon_1\eta(x) + \epsilon_2\zeta(x)$, where ϵ_1 and ϵ_2 are parameters and $\eta(x)$ and $\zeta(x)$ are arbitrary functions satisfying the conditions $\eta(x_0) = \eta(x_1) = \zeta(x_0) = \zeta(x_1) = 0$. Then the function

$$\Phi(\epsilon_1, \epsilon_2) = \int_{x_0}^{x_1} F(x, y + \epsilon_1\eta + \epsilon_2\zeta, y' + \epsilon_1\eta' + \epsilon_2\zeta')\, dx$$

must be stationary at $\epsilon_1 = \epsilon_2 = 0$ with respect to all sufficiently small values of ϵ_1 and ϵ_2 for which

$$\Psi(\epsilon_1, \epsilon_2) = \int_{x_0}^{x_1} G(x, y + \epsilon_1\eta + \epsilon_2\zeta, y' + \epsilon_1\eta' + \epsilon_2\zeta')\, dx = c.$$

According to the theorems on ordinary maxima and minima (see §1) there exist two constants λ_0 and λ, not both equal to zero, such that

$$\frac{\partial}{\partial\epsilon_1}[\lambda_0\Phi(\epsilon_1, \epsilon_2) + \lambda\Psi(\epsilon_1, \epsilon_2)]|_{\epsilon_1=\epsilon_2=0} = 0,$$

$$\frac{\partial}{\partial\epsilon_2}[\lambda_0\Phi(\epsilon_1, \epsilon_2) + \lambda\Psi(\epsilon_1, \epsilon_2)]|_{\epsilon_1=\epsilon_2=0} = 0.$$

We thus have

$$\int_{x_0}^{x_1} \{\lambda_0[F]_y + \lambda[G]_y\}\eta\, dx = 0,$$

$$\int_{x_0}^{x_1} \{\lambda_0[F]_y + \lambda[G]_y\}\zeta\, dx = 0.$$

We may conclude from the first of these equations that the ratio of λ_0 to λ does not depend on ζ. Since ζ is arbitrary, it follows from the second equation that $\lambda_0[F]_y + \lambda[G]_y = 0$. If $\lambda_0 \neq 0$, i.e. if equation[1]

$$(G_{y'})' - G_y = 0 \tag{52}$$

[1] It is easily seen that we have the exceptional case (52) whenever there exists only one function which satisfies the given subsidiary condition.

does not hold, we may set $\lambda_0 = 1$, and we have

$$\frac{d}{dx}\frac{\partial(F + \lambda G)}{\partial y'} - \frac{\partial(F + \lambda G)}{\partial y} = 0. \tag{53}$$

Thus we have the result: *Except in the exceptional case where equation* (52) *holds, the Euler equation of our variational problem is obtained by forming the Euler equation of the variational problem with integrand* $F^* = F + \lambda G$ *for a suitable value of the parameter* λ, *disregarding the subsidiary condition.*

The general integral of the differential equation (53) contains the parameter λ in addition to the two constants of integration. These three quantities must be determined from the boundary conditions and the equation $K = c$.

A simple example is given by the ordinary isoperimetric problem in which $F = \sqrt{1 + y'^2}$ and $G = y$. We obtain immediately

$$\frac{d}{dx}\frac{\partial}{\partial y'}(\sqrt{1 + y'^2} + \lambda y) - \frac{\partial}{\partial y}(\sqrt{1 + y'^2} + \lambda y) = 0$$

or

$$\frac{d}{dx}\frac{y'}{\sqrt{1 + y'^2}} = \lambda,$$

from which we find that the extremals are circles.

Another example concerns the equilibrium position of a homogeneous string suspended at its end points. Here $F = y\sqrt{1 + y'^2}$ and $G = \sqrt{1 + y'^2}$; we obtain (using the Euler equation for $F_x = 0$, page 206)

$$F^* - y'F^*_{y'} = (y + \lambda)\left(\sqrt{1 + y'^2} - \frac{y'^2}{\sqrt{1 + y'^2}}\right) = \frac{y + \lambda}{\sqrt{1 + y'^2}} = c,$$

$$y + \lambda = c \cosh\left(\frac{x}{c} + c_1\right);$$

thus the required curve is a *catenary*. The exceptional case may be illustrated by the subsidiary condition $\int_0^1 \sqrt{1 + y'^2}\, dx = 1$, with $y(0) = y(1) = 0$. Evidently $y = 0$ is the only admissible comparison function, and it does in fact satisfy equation (52). Whatever F may be, the solution cannot be anything but $y = 0$.[1]

[1] For a study of the exceptional case see C. Carathéodory, Über die diskontinuierlichen Lösungen in der Variationsrechnung, Göttingen, 1904, pp. 45 ff.

2. Finite Subsidiary Conditions. We shall now consider another type of variational problem: Make the integral $J = \int_{x_0}^{x_1} F(x, y, z, y', z')\,dx$ stationary in comparison to functions $y(x)$, $z(x)$ which satisfy, in addition to the boundary conditions $y(x_0) = y_0$, $y(x_1) = y_1$, $z(x_0) = z_0$, $z(x_1) = z_1$, a subsidiary condition of the form

$$G(x, y, z) = 0. \tag{54}$$

Geometrically speaking, a space curve $y(x)$, $z(x)$ lying on a given surface is to be determined by the extremum requirement.[1]

A natural way to obtain necessary conditions for the functions $y(x)$ and $z(x)$ is to solve the equation $G(x, y, z) = 0$ for one of the functions, say $z(x)$, thus reducing the problem to that of determining *one* independent function $y(x)$. According to elementary theorems of analysis this solution $z = g(x, y)$ can certainly be obtained if $\partial G/\partial z \neq 0$ on the extremal in question. We may then regard z' as a function of x, y, y' and thus eliminate it from $F(x, y, z, y', z')$ by noting the relation $G_x + y'G_y + z'G_z = 0$ or $z' = y'\partial g/\partial y + \partial g/\partial x$. Thus we have

$$F(x, y, z, y', z') = F\left(x, y, g(x, y), y', \frac{\partial g}{\partial x} + y'\frac{\partial g}{\partial y}\right)$$

and y must satisfy the Euler equation

$$\frac{d}{dx}\left(F_{y'} + F_{z'}\frac{\partial g}{\partial y}\right) - \left[F_y + F_z\frac{\partial g}{\partial y} + F_{z'}\left(\frac{\partial^2 g}{\partial x\,\partial y} + y'\frac{\partial^2 g}{\partial y^2}\right)\right] = 0,$$

which is easily transformed to the form

$$(F'_{y'} - F_y) + (F'_{z'} - F_z)\frac{\partial g}{\partial y} = 0.$$

But, since

$$G_y + G_z\frac{\partial g}{\partial y} = 0,$$

the proportion

$$(F'_{y'} - F_y) : (F'_{z'} - F_z) = G_y : G_z$$

must hold. Therefore either $G_y = G_z = 0$ identically along the ex-

[1] Note that the coordinate x is singled out, so that not all curves on the surface $G = 0$ are necessarily admissible.

tremal (which contradicts our hypothesis) or there exists a factor of proportionality $\lambda = \lambda(x)$ for which

$$F'_{y'} - F_y - \lambda G_y = 0; \qquad F'_{z'} - F_z - \lambda G_z = 0. \tag{55}$$

If we set $F^* = F + \lambda G$ the result may be written in the form of the Euler equations for F^*:

$$-[F^*]_y = F^{*\prime}_{y'} - F^*_y = 0; \qquad -[F^*]_z = F^{*\prime}_{z'} - F^*_z = 0.$$

These equations are necessary conditions for an extremum unless the two equations

$$G_y = 0, \qquad G_z = 0$$

hold simultaneously on the extremal; in the latter case the third equation $G_x = 0$ holds because of the relation $G_x + y'G_y + z'G_z = 0$.

The factors λ which occur here and in the preceding example are known as *Euler* or *Lagrange multipliers*; analogous terminology is used in differential calculus. *A formal similarity exists between the problems discussed in subsections 1 and 2: the expression $F^* = F + \lambda G$ is formed and the Euler equation for F^* is set up in both cases. However, in the first case λ is a constant while in the second case it is a function of x. The Euler equations, together with the subsidiary condition and the boundary conditions, supply the correct number of conditions for determining the extremal.*

A special case of the problem just discussed is to determine the *geodesics on a given surface* $G(x, y, z) = 0$. Here $F = \sqrt{1 + y'^2 + z'^2}$, and we obtain for the geodesic curves, represented parametrically by $x = x(t)$, $y = y(t)$, $z = z(t)$,

$$\frac{d}{dt}\frac{\dot{x}}{\sqrt{\dot{x}^2 + \dot{y}^2 + \dot{z}^2}} : \frac{d}{dt}\frac{\dot{y}}{\sqrt{\dot{x}^2 + \dot{y}^2 + \dot{z}^2}} : \frac{d}{dt}\frac{\dot{z}}{\sqrt{\dot{x}^2 + \dot{y}^2 + \dot{z}^2}} = G_x : G_y : G_z$$

or

$$\frac{d}{dt}\frac{\dot{x}}{\sqrt{\dot{x}^2 + \dot{y}^2 + \dot{z}^2}} - \lambda G_x = 0,$$

$$\frac{d}{dt}\frac{\dot{y}}{\sqrt{\dot{x}^2 + \dot{y}^2 + \dot{z}^2}} - \lambda G_y = 0,$$

$$\frac{d}{dt}\frac{\dot{z}}{\sqrt{\dot{x}^2 + \dot{y}^2 + \dot{z}^2}} - \lambda G_z = 0.$$

These three equations, together with the fourth equation $G = 0$, determine the geodesic curves and the multiplier $\lambda(x)$. This representation brings out the most important geometrical properties of geodesics; for example, it shows that the osculating plane goes through the normal to the surface. The proof is left to the reader.

3. Differential Equations as Subsidiary Conditions. Up to now the multiplier λ has been used merely as an elegant artifice. But multipliers are indispensable if the subsidiary condition takes the general form

$$G(x, y, z, y', z') = 0, \tag{56}$$

where the expression $G(x, y, z, y', z')$ cannot be obtained by differentiating an expression $H(x, y, z)$ with respect to x, i.e. where G is a *nonintegrable differential expression*. Such subsidiary conditions are also called *nonholonomic* conditions. A simple example of such a condition is $y' - z = 0$. If this condition were holonomic, i.e. equivalent to a finite condition $H(x, y, z) = \text{const.}$, the values of x, y, and z could not be chosen independently everywhere, but for every set of values x, y, z it is evidently possible to choose y' so that it satisfies the condition $y' - z = 0$. Nonholonomic conditions occur in mechanics when the equations of constraint contain direction as well as position coordinates, such as in the motion of a ship, a skate, or a rolling sphere.

The problems with subsidiary conditions already treated may be regarded as special cases of this general problem. This is obvious for the problem of subsection 2. But the isoperimetric problem may also be brought into this category. In this case z and z' do not occur at all in F, while the subsidiary condition takes the form $z' - G(x, y, y') = 0$. The boundary conditions are

$$y(x_0) = y_0\,, \qquad y(x_1) = y_1\,, \qquad z(x_0) = 0, \qquad z(x_1) = c.$$

The ordinary minimum problem with higher derivatives under the integral sign is also a special case of this problem. For example, the problem of the extremum of $\int_{x_0}^{x_1} F(x, y, y', y'')\,dx$ is equivalent to that of the extremum of $\int_{x_0}^{x_1} F(x, y, y', z')\,dx$ with the subsidiary condition $z - y' = 0$.

In all these special cases the necessary conditions can obviously be expressed as follows: *If the solution does not satisfy the Euler equations*

corresponding to the expression G, then there exists a multiplier $\lambda(x)$ such that the Euler equations corresponding to the expression $F^* = F + \lambda G$ are satisfied.

This multiplier rule also holds for the general problem formulated above. We shall omit the proof, referring the reader to the literature.[1]

In conclusion, it should be emphasized that our method remains valid if the number of unknown functions and subsidiary conditions is increased. For functions of several independent variables, our results are presumably valid, although a general proof has not yet been given in the case where the subsidiary conditions are partial differential equations.

§8. *Invariant Character of the Euler Equations*

1. The Euler Expression as a Gradient in Function Space. Invariance of the Euler Expression. The stationary character of a function $f(x_1, x_2, \cdots, x_n)$ at a particular point is equivalent to the equation

$$\text{grad } f = 0,$$

where grad f denotes the gradient of the function, i.e. the vector in n-dimensional space whose components are $f_{x_1}, f_{x_2}, \cdots, f_{x_n}$. This gradient vector has the following characteristic property: if the n variables $x_1, x_2, \cdots, x_n$ are all differentiable functions of a parameter t, then $f(x_1, x_2, \cdots, x_n)$ goes over into a function of t, and we have

$$\dot{f}(t) = \sum_{i=1}^{n} \dot{x}_i f_{x_i} = \mathbf{v} \cdot \text{grad } f, \tag{57}$$

where the dot denotes differentiation with respect to t and $\mathbf{v}$ is the "*translation vector*" with the components $\dot{x}_i$; thus the rate of change of the function is the inner product of the translation vector of the argument point and the gradient vector.

The Euler differential expression, which vanishes if and only if the functional is stationary, may be regarded analogously as the *gradient of a functional in function space.*

For example, in the case of a functional

$$J[\varphi] = \int_{x_0}^{x_1} F(x, \varphi, \varphi')\, dx$$

[1] See D. Hilbert, Zur Variationsrechnung, Math. Ann., Vol. 62, 1906, pp. 351–370. A detailed exposition is also found in the texts by Bolza and Hadamard referred to in the bibliography.

we may suppose that the argument function φ depends on a parameter t as well as on the independent variable x. Then $J[\varphi] = J(t)$ is a function of t, and from the expression for the first variation we obtain

$$\dot{J}(t) = \int_{x_0}^{x_1} \dot{\varphi}(x)[F]_\varphi \, dx$$

if, independent of t, φ remains fixed at the boundary of the interval. Here the dot again denotes differentiation with respect to t. This formula is completely analogous to formula (57) for the function $f(x_1, x_2, \cdots, x_n)$. We express this analogy by calling the expression $[F]_\varphi$ the gradient of $J[\varphi]$ in function space.

In general one may define the gradient of a functional $J[\varphi]$ as an expression $G[\varphi]$ such that, if φ is replaced by a family of functions depending on a parameter t and the independent variables, the relation

$$\frac{d}{dt} J[\varphi] = \int_{x_0}^{x_1} \dot{\varphi} \, G[\varphi] \, dx$$

holds.

For example, if $K(x, y) = K(y, x)$, the expression $2\int_0^1 K(x, y)\varphi(y)\, dy$ is the gradient of the functional $\int_0^1 \int_0^1 K(x, y)\varphi(x)\varphi(y)\, dx\, dy$.

The gradient of a function has some well-known invariance properties under transformations of the independent variables. The Euler differential expression has analogous invariant (or rather covariant) properties when the functions in the integrand are transformed to new independent variables.

In the simplest case let us transform x to the new variable $\xi(x)$ and let

$$F(x, y, y') = F\left(x(\xi), y, \frac{dy/d\xi}{dx/d\xi}\right) = \Phi\left(\xi, y, \frac{dy}{d\xi}\right),$$

so that $\int_{x_0}^{x_1} F\, dx = \int_{\xi_0}^{\xi_1} \Phi \frac{dx}{d\xi}\, d\xi$. Then

$$\begin{aligned} \int_{x_0}^{x_1} [F]_\eta \, dx &= \frac{\partial}{\partial \epsilon} \int_{x_0}^{x_1} F(x, y + \epsilon\eta, y' + \epsilon\eta')\, dx \Big|_{\epsilon=0} \\ &= \frac{\partial}{\partial \epsilon} \int_{\xi_0}^{\xi_1} \Phi\left(\xi, y + \epsilon\eta, \frac{dy}{d\xi} + \epsilon \frac{d\eta}{d\xi}\right) \frac{dx}{d\xi}\, d\xi \Big|_{\epsilon=0} \\ &= \int_{\xi_0}^{\xi_1} \left[\Phi \frac{dx}{d\xi}\right]_y \eta \, d\xi; \end{aligned} \tag{58}$$

therefore, since η is arbitrary (except that it must vanish at the boundary), we have

$$[F]_u = \frac{d\xi}{dx}\left[\Phi\,\frac{dx}{d\xi}\right]_v \tag{59}$$

In the case of two independent variables we obtain in the same way

$$\begin{aligned} F(x, y, u, u_x, u_y) &= F(x(\xi, \eta), y(\xi, \eta), u, u_\xi\xi_x + u_\eta\eta_x, u_\xi\xi_y + u_\eta\eta_y) \\ &= \Phi(\xi, \eta, u, u_\xi, u_\eta); \end{aligned}$$

$$\iint_G F\,dx\,dy = \iint_G \Phi\,\frac{\partial(x, y)}{\partial(\xi, \eta)}\,d\xi\,d\eta,$$

$$\iint_G [F]_u\zeta\,dx\,dy = \iint_G \left[\Phi\,\frac{\partial(x, y)}{\partial(\xi, \eta)}\right]_u \zeta\,d\xi\,d\eta,$$

$$[F]_u = \frac{\partial(\xi, \eta)}{\partial(x, y)}\left[\Phi\,\frac{\partial(x, y)}{\partial(\xi, \eta)}\right]_u. \tag{60}$$

Analogous transformation properties hold when there are more than two independent variables.

The invariance property, elegantly expressed by these formulas, is of great practical advantage. It simplifies computations involved in transforming Euler differential expressions to new independent variables, since the transformation of the second derivatives need not be carried out explicitly.

2. Transformation of Δu. Spherical coordinates. One important example is the integrand $u_x^2 + u_y^2 + u_z^2$. Suppose that the transformation $x = x(\xi_1, \xi_2, \xi_3)$, $y = y(\xi_1, \xi_2, \xi_3)$, $z = z(\xi_1, \xi_2, \xi_3)$ changes the square of the line element $dx^2 + dy^2 + dz^2$ into $\sum_{i,k} g_{ik}\,d\xi_i\,d\xi_k$, where

$$g_{ik} = \frac{\partial x}{\partial \xi_i}\frac{\partial x}{\partial \xi_k} + \frac{\partial y}{\partial \xi_i}\frac{\partial y}{\partial \xi_k} + \frac{\partial z}{\partial \xi_i}\frac{\partial z}{\partial \xi_k};$$

the determinant $a = |g_{ik}|$ of these quantities is the square of the Jacobian determinant of x, y, z with respect to ξ_1, ξ_2, ξ_3. It is easily seen that

$$u_x^2 + u_y^2 + u_z^2 = \sum_{i,k} g^{ik}u_i\,u_k \qquad \left(u_i = \frac{\partial u}{\partial \xi_i}\right),$$

where the quantities g^{ik} are defined by

$$g^{ik} = \frac{\partial \xi_i}{\partial x}\frac{\partial \xi_k}{\partial x} + \frac{\partial \xi_i}{\partial y}\frac{\partial \xi_k}{\partial y} + \frac{\partial \xi_i}{\partial z}\frac{\partial \xi_k}{\partial z}$$

and satisfy the equations

$$\sum_i g_{ik} g^{il} = \delta_{kl},$$

δ_{kl} being the Kronecker delta, $\delta_{kl} = 0$ for $k \neq l$, $\delta_{kk} = 1$.

We therefore obtain the following general equation for the transformation of Δu to curvilinear coordinates ξ_1, ξ_2, ξ_3:

$$\Delta u = \frac{1}{\sqrt{a}} \sum_i \frac{\partial}{\partial \xi_i}\left(\sqrt{a} \sum_k g^{ik} u_k\right). \tag{61}$$

In particular, if $g_{12} = g_{13} = g_{23} = 0$, i.e. if the new coordinate system also is orthogonal (the coordinate surfaces $\xi_1 = \text{const.}$, $\xi_2 = \text{const.}$, $\xi_3 = \text{const.}$ intersect orthogonally), the transformation formula becomes

$$\Delta u = \frac{\dfrac{\partial}{\partial \xi_1}\left(u_1 \sqrt{\dfrac{g_{22} g_{33}}{g_{11}}}\right) + \dfrac{\partial}{\partial \xi_2}\left(u_2 \sqrt{\dfrac{g_{33} g_{11}}{g_{22}}}\right) + \dfrac{\partial}{\partial \xi_3}\left(u_3 \sqrt{\dfrac{g_{11} g_{22}}{g_{33}}}\right)}{\sqrt{g_{11} g_{22} g_{33}}}. \tag{62}$$

For example, for *spherical coordinates* r, θ, φ,

$$x = r \sin\theta \cos\varphi, \qquad y = r \sin\theta \sin\varphi, \qquad z = r \cos\theta,$$

$$ds^2 = dr^2 + r^2\, d\theta^2 + r^2 \sin^2\theta\, d\varphi^2,$$

we obtain after a brief calculation

$$\Delta u = \frac{1}{r^2 \sin\theta}\left\{\frac{\partial}{\partial r}(r^2 u_r \sin\theta) + \frac{\partial}{\partial \theta}(u_\theta \sin\theta) + \frac{\partial}{\partial \varphi}\left(\frac{u_\varphi}{\sin\theta}\right)\right\}. \tag{63}$$

In the case of only two independent variables ξ, η, the corresponding formulas apply. Thus if

$$ds^2 = e\, d\xi^2 + 2f\, d\xi\, d\eta + g\, d\eta^2$$

we obtain the invariant form of the differential expression

$$\Delta u = \frac{1}{\sqrt{eg - f^2}}\left\{\frac{\partial}{\partial \xi}\left(\frac{g u_\xi - f u_\eta}{\sqrt{eg - f^2}}\right) + \frac{\partial}{\partial \eta}\left(\frac{e u_\eta - f u_\xi}{\sqrt{eg - f^2}}\right)\right\}. \tag{64}$$

In particular, in the case of polar coordinates,

$$(65) \qquad ds^2 = dr^2 + r^2\, d\varphi^2, \qquad \Delta u = \frac{1}{r}\left\{\frac{\partial}{\partial r}(ru_r) + \frac{\partial}{\partial \varphi}\left(\frac{u_\varphi}{r}\right)\right\}.$$

3. Ellipsoidal Coordinates.[1] The transformation to ellipsoidal coordinates is also important. These coordinates are defined as the three roots ρ, σ, τ of the cubic equation in s

$$(66) \qquad \frac{x^2}{s - e_1} + \frac{y^2}{s - e_2} + \frac{z^2}{s - e_3} = 1,$$

where e_1, e_2, e_3 are given real numbers. These roots are real for real x, y, z and, if $e_1 > e_2 > e_3$, they can be so labelled as to satisfy the inequalities

$$\rho \geq e_1 \geq \sigma \geq e_2 \geq \tau \geq e_3 .$$

The surfaces $\rho = \text{const.}$, $\sigma = \text{const.}$, $\tau = \text{const.}$ are ellipsoids, hyperboloids of one sheet, and hyperboloids of two sheets, respectively. Cartesian coordinates are expressed in terms of ellipsoidal coordinates by

$$(67) \qquad \begin{aligned} x^2 &= \frac{(\rho - e_1)(\sigma - e_1)(\tau - e_1)}{(e_1 - e_2)(e_1 - e_3)}, \\ y^2 &= \frac{(\rho - e_2)(\sigma - e_2)(\tau - e_2)}{(e_2 - e_3)(e_2 - e_1)}, \\ z^2 &= \frac{(\rho - e_3)(\sigma - e_3)(\tau - e_3)}{(e_3 - e_1)(e_3 - e_2)} \end{aligned}$$

and the line element is

$$(68) \qquad \begin{aligned} 4\, ds^2 = {} & \frac{(\rho - \sigma)(\rho - \tau)}{(\rho - e_1)(\rho - e_2)(\rho - e_3)}\, d\rho^2 \\ & + \frac{(\sigma - \tau)(\sigma - \rho)}{(\sigma - e_1)(\sigma - e_2)(\sigma - e_2)}\, d\sigma^2 \\ & + \frac{(\tau - \sigma)(\tau - \rho)}{(\tau - e_1)(\tau - e_2)(\tau - e_3)}\, d\tau^2. \end{aligned}$$

[1] See C. G. J. Jacobi, Vorlesungen über Dynamik (given at Königsberg 1842–3, published by A. Clebsch, Berlin, 1866, reprinted as a supplement to Jacobi's Collected Works, Berlin, 1884), Lecture 26, where the details of the calculation may be found. It is to be emphasized that the following considerations may be immediately generalized to more than three dimensions.

This suggests the introduction of the new variables

$$t_1 = \int^{\rho} \frac{d\lambda}{\sqrt{f(\lambda)}}, \qquad t_2 = \int^{\sigma} \frac{d\lambda}{\sqrt{f(\lambda)}}, \qquad t_3 = \int^{\tau} \frac{d\lambda}{\sqrt{f(\lambda)}}$$

$$\text{with} \quad f(\lambda) = 4(\lambda - e_1)(\lambda - e_2)(\lambda - e_3).$$

If $e_1 + e_2 + e_3 = 0$, which can be achieved by the substitution $s = s' + \frac{1}{3}(e_1 + e_2 + e_3)$, and if the lower limit of the integrals is taken to be ∞, we obtain simply

$$\rho = \mathfrak{p}(t_1), \qquad \sigma = \mathfrak{p}(t_2), \qquad \tau = \mathfrak{p}(t_3),$$

where $\mathfrak{p}$ is Weierstrass's $\mathfrak{p}$-function;[1] furthermore

$$ds^2 = (\rho - \sigma)(\rho - \tau)\, dt_1^2 + (\sigma - \tau)(\sigma - \rho)\, dt_2^2 + (\tau - \rho)(\tau - \sigma)\, dt_3^2 .$$

For a function T of the coordinates t_i we have, in accordance with (62),

$$\Delta T = \frac{\dfrac{\partial}{\partial t_1}\left((\tau - \sigma)\frac{\partial T}{\partial t_1}\right) + \dfrac{\partial}{\partial t_2}\left((\rho - \tau)\frac{\partial T}{\partial t_2}\right) + \dfrac{\partial}{\partial t_3}\left((\sigma - \rho)\frac{\partial T}{\partial t_3}\right)}{(\sigma - \tau)(\tau - \rho)(\rho - \sigma)}$$

(69)

$$= \frac{1}{(\rho - \sigma)(\rho - \tau)} \frac{\partial^2 T}{\partial t_1^2} + \frac{1}{(\sigma - \tau)(\sigma - \rho)} \frac{\partial^2 T}{\partial t_2^2} + \frac{1}{(\tau - \rho)(\tau - \sigma)} \frac{\partial^2 T}{\partial t_3^2} .$$

One advantage of the introduction of the integrals t_i is that the Cartesian coordinates are single-valued functions of the t_i since the expressions in the numerators of

$$x = \frac{\sqrt{\mathfrak{p}(t_1) - e_1}\,\sqrt{\mathfrak{p}(t_2) - e_1}\,\sqrt{\mathfrak{p}(t_3) - e_1}}{\sqrt{e_1 - e_2}\,\sqrt{e_1 - e_3}},$$

(70)

$$y = \frac{\sqrt{\mathfrak{p}(t_1) - e_2}\,\sqrt{\mathfrak{p}(t_2) - e_2}\,\sqrt{\mathfrak{p}(t_3) - e_2}}{\sqrt{e_2 - e_3}\,\sqrt{e_2 - e_1}},$$

$$z = \frac{\sqrt{\mathfrak{p}(t_1) - e_3}\,\sqrt{\mathfrak{p}(t_2) - e_3}\,\sqrt{\mathfrak{p}(t_3) - e_3}}{\sqrt{e_3 - e_1}\,\sqrt{e_3 - e_2}}$$

[1] See, for example, Hurwitz and Courant, Vorlesungen über allgemeine Funktionentheorie und elliptische Funktionen, pp. 161–171, 3rd ed., Springer, Berlin, 1929; Interscience, New York, 1944.

are single-valued functions of t_1, t_2, t_3 once the signs of the square roots are fixed. As the point (x, y, z) ranges over an octant, each of the quantities ρ, σ, τ traverses its appropriate interval, and when one of these assumes one of its end values, the point (x, y, z) lies on a boundary plane of the octant as indicated in Figure 1. In this dia-

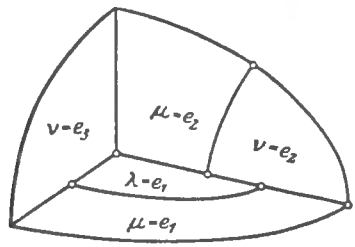


Figure 1. Confocal surfaces of the second order.

gram, the planes are cut off at their intersections with an ellipsoid $\rho = \rho_1 > e_1$; the inner boundary curves are portions of the

$$\text{"}\textit{focal ellipse}\text{"} \quad \left(x = 0, \qquad \frac{y^2}{e_1 - e_2} + \frac{z^2}{e_1 - e_3} = 1\right)$$

and of the

$$\text{"}\textit{focal hyperbola}\text{"} \quad \left(y = 0, \qquad \frac{x^2}{e_2 - e_1} + \frac{z^2}{e_2 - e_3} = 1\right).$$

Now if ω and ω' are, respectively, the real and pure imaginary periods of the integrals t_i, i.e. if

$$\omega = 2\int_{e_1}^{\infty} \frac{d\lambda}{\sqrt{f(\lambda)}}, \qquad \omega' = 2\int_{-\infty}^{e_3} \frac{d\lambda}{\sqrt{f(\lambda)}},$$

we may let t_1 vary from 0 to $\omega/2$, t_2 from $\omega/2$ to $\frac{1}{2}(\omega + \omega')$, and t_3 from $\frac{1}{2}(\omega + \omega')$ to $\frac{1}{2}\omega'$, obtaining all the points of the octant. If the interval for each t_i is doubled the point ranges over the whole space. If a single-valued function of the t_i is to be single-valued in space, it must remain unchanged under all substitutions of the t_i which leave x, y, and z unchanged, for example it is single-valued if t_1 and t_2 are replaced by $\omega - t_1$ and $\omega - t_2$, respectively.

If we write $t_1 = u$, $t_2 = \omega/2 + iv$, $t_3 = \omega'/2 + w$, $\mathfrak{p}(t_1) = f(u)$, $\mathfrak{p}(t_2) =$

$g(v)$, $\mathfrak{p}(t_3) = h(w)$, we may take u, v, and w to be real. Then

$$
\begin{aligned}
ds^2 &= [f(u) - g(v)][f(u) - h(w)]\,du^2 \\
&+ [f(u) - g(v)][g(v) - h(w)]\,dv^2 \\
&+ [f(u) - h(w)][g(v) - h(w)]\,dw^2
\end{aligned} \tag{71}
$$

and for real u, v, w all the coefficients are non-negative, since $f(u) \geq e_1 \geq g(v) \geq e_2 \geq h(w) \geq e_3$. The fact that dt_2 is pure imaginary in the symmetric form in t_1, t_2, t_3 is important; for, the positive definite character of ds^2 is assured by the negative value of the coefficient of dt_2^2.

Among the *degenerate forms of ellipsoidal coordinates* we may mention (aside from spherical coordinates, which may also be regarded as a degenerate case) the *spheroidal* and *paraboloidal* coordinates. If two of the e_i, say e_1 and e_2, coincide we obtain

$$
\frac{x^2 + y^2}{s - e_1} + \frac{z^2}{s - e_3} = 1. \tag{72}
$$

The two roots $s = \lambda_1$, $s = \lambda_2$ of this equation, together with the angle φ defined by

$$
x = r\cos\varphi, \qquad y = r\sin\varphi, \qquad r^2 = x^2 + y^2,
$$

form the new coordinates. We have here

$$
r^2 = \frac{(\lambda_1 - e_1)(\lambda_2 - e_1)}{e_3 - e_1}, \qquad z^2 = \frac{(\lambda_1 - e_3)(\lambda_2 - e_3)}{e_1 - e_3}, \tag{73}
$$

$$
\begin{aligned}
ds^2 &= r^2\,d\varphi^2 + \frac{\lambda_1 - \lambda_2}{4(\lambda_1 - e_1)(\lambda_1 - e_3)}\,d\lambda_1^2 + \frac{\lambda_2 - \lambda_1}{4(\lambda_2 - e_1)(\lambda_2 - e_3)}\,d\lambda_2^2 \\
&= r^2\,d\varphi^2 + (\lambda_1 - \lambda_2)(dt_1^2 - dt_2^2)
\end{aligned} \tag{74}
$$

with

$$
t_i = \int^{\lambda_i} \frac{d\lambda}{\sqrt{4(\lambda - e_1)(\lambda - e_3)}}. \tag{75}
$$

Hence

$$
\Delta T = \frac{1}{r^2}\frac{\partial^2 T}{\partial \varphi^2} + \frac{1}{r(\lambda_1 - \lambda_2)}\left[\frac{\partial}{\partial t_1}\left(r\frac{\partial T}{\partial t_1}\right) - \frac{\partial}{\partial t_2}\left(r\frac{\partial T}{\partial t_2}\right)\right]. \tag{76}
$$

If we now let one end of the ellipsoids tend to infinity we obtain,

after passage to the limit,[1] the *paraboloidal* coordinates as the roots λ_1, λ_2 of equation

$$\frac{x^2 + y^2}{s - e_1} - 2z + s - e_1 = 0, \tag{77}$$

where r and z are given by the following expressions:

$$r^2 = -(\lambda_1 - e_1)(\lambda_2 - e_1), \qquad 2z = 2e_1 - \lambda_1 - \lambda_2 . \tag{78}$$

Here the coordinates of a point in space are λ_1, λ_2, and φ. The line element is (74)

$$\begin{aligned} ds^2 &= r^2\, d\varphi^2 + \frac{\lambda_1 - \lambda_2}{4(\lambda_1 - e_1)}\, d\lambda_1^2 + \frac{\lambda_2 - \lambda_1}{4(\lambda_2 - e_1)}\, d\lambda_2^2 \\ &= r^2\, d\varphi^2 + (\lambda_1 - \lambda_2)(dt_1^2 - dt_2^2) \end{aligned}$$

with

$$t_i = \int^{\lambda_i} \frac{d\lambda}{\sqrt{4(\lambda - e_1)}} = \sqrt{\lambda_i - e_1}, \tag{79}$$

and the differential expression ΔT takes the form (76)

$$\Delta T = \frac{1}{r^2}\frac{\partial^2 T}{\partial \varphi^2} + \frac{1}{r(\lambda_1 - \lambda_2)}\left[\frac{\partial}{\partial t_1}\left(r\frac{\partial T}{\partial t_1}\right) - \frac{\partial}{\partial t_2}\left(r\frac{\partial T}{\partial t_2}\right)\right].$$

If in the above expressions the terms containing φ are omitted, one is immediately led to the formulas for elliptic and parabolic coordinates in the r, z-plane. In both cases we obtain, from formula (64),

$$\Delta T = \frac{1}{\lambda_1 - \lambda_2}\left(\frac{\partial^2 T}{\partial t_1^2} - \frac{\partial^2 T}{\partial t_2^2}\right),$$

in which λ_i and t_i are connected by

$$t_i = \int^{\lambda_i} \frac{d\lambda}{\sqrt{4(\lambda - e_1)(\lambda - e_3)}}$$

and

$$t_i = \sqrt{\lambda_i - e_1},$$

respectively.

[1] These coordinates may, of course, also be defined without reference to the foregoing treatment by starting out from the system (77) of confocal paraboloids of revolution.

§9. *Transformation of Variational Problems to Canonical and Involutory Form*

The Lagrange multiplier method leads to several transformations which are important both theoretically and practically.

By means of these transformations new problems equivalent to a given problem can be so formulated that stationary conditions occur simultaneously in equivalent problems. In this way we are led to transformations of the variational problems which are important because of their symmetric character. Moreover, for a given minimum problem with minimum d, we shall often be able to find an equivalent maximum problem with the same value d as maximum; this is a useful tool for the practical problem of bounding d from above and below.[1]

1. Transformation of an Ordinary Minimum Problem with Subsidiary Conditions. Before discussing these transformations, we briefly consider ordinary minimum problems with a finite number of variables. Our discussion is based on the following self-evident principle: *If a function* $f(x_1, x_2, \cdots, x_n)$, *subject to certain subsidiary conditions, has a stationary value at the point* $x_i = \xi_i \quad (i = 1, 2, \cdots, n)$ *and if the quantities* ξ_i *satisfy any relation* $r(\xi_1, \xi_2, \cdots, \xi_n) = 0$, *then* f *remains stationary at* $x_i = \xi_i$ *provided the additional condition* $r(x_1, x_2, \cdots, x_n) = 0$ *is appended to the subsidiary conditions.*

We begin by considering the problem

I: $f(x, y)$ *is to be made stationary under the subsidiary condition* $g(x, y) = 0$; *the ordinary continuity and differentiability requirements are to be satisfied, and we suppose that* $g_x^2 + g_y^2 \neq 0$ *at the stationary point.* By the multiplier rule, problem I may be replaced by the equivalent problem

II: $F(x, y; \lambda) = f(x, y) + \lambda g(x, y)$ *is to be made stationary as a function of the three arguments* x, y, λ.

The condition $dF = 0$ is then equivalent to the three equations $f_x + \lambda g_x = 0$, $f_y + \lambda g_y = 0$, $g = 0$. Had we started with problem II, then by adding explicitly the condition $g = 0$, which is automatically fulfilled by the solution of problem II, we would immediately have arrived at problem I by our general principle.

[1] See E. Trefftz, Ein Gegenstück zum Ritzschen Verfahren, Verh. d. 2. Int. Kongr. für Technische Mechanik, Zürich, 1927, p. 131, where such an approximation procedure was first given. See also Trefftz, Konvergenz und Fehlerschätzung beim Ritzschen Verfahren, Math. Ann., Vol. 100, 1928, pp. 503–521.

But we may also obtain another problem equivalent to problem II (by equivalence we mean that stationary behavior occurs at the same point) by appending as subsidiary conditions, not $g = 0$, but the other two equations which the solution of II satisfies. We thus arrive at problem

III: $F(x, y; \lambda) = f + \lambda g$ *is to be made stationary under the subsidiary conditions* $f_x + \lambda g_x = 0, f_y + \lambda g_y = 0$.

If we assume that the latter two equations can be solved, in the vicinity of the stationary point, for x and y as functions of λ, then $F(x, y; \lambda)$ goes over into a function $\psi(\lambda)$ of λ and we obtain problem IV, which once again is equivalent to the other three problems:

IV: $\psi(\lambda)$ *is to be made stationary.*

We shall now investigate the stationary points; are they or are they not maxima or minima? Let us suppose in problem I, henceforth denoted by I′, that f possesses an actual minimum $f(\bar{x}, \bar{y}) = d$ at the point $\bar{x}, \bar{y}$. We then consider problem

II′: $F(x, y; \lambda) = f + \lambda g = min.$ *with fixed* λ. Let us assume that, if λ is chosen arbitrarily in a certain neighborhood of the value $\bar{\lambda}$ defined by the Lagrange multiplier rule, there exists an actual minimum, which we denote by $d_\lambda = \psi(\lambda)$, and which is characterized by the equations $f_x + \lambda g_x = 0, f_y + \lambda g_y = 0$. Then we certainly have

$$d_\lambda \leq d.$$

Indeed, problem I′ with the minimum d is obtained from problem II′ with minimum d_λ by appending the condition $g = 0$, which restricts the domain of comparison values. If we further assume that for every λ in the neighborhood of $\bar{\lambda}$ the equations $f_x + \lambda g_x = 0$, $f_y + \lambda g_y = 0$ *uniquely* determine x and y as functions of λ, then $d_{\bar{\lambda}} = d$, and

$$d = \text{max. } (d_\lambda).$$

Thus d is the maximum of the minimum $\psi(\lambda)$ of $F = f + \lambda g$ where the minimum is to be taken for fixed λ and then the maximum taken with respect to λ. Under these conditions we may also characterize d by the problem

III′: $F(x, y; \lambda) = f + \lambda g = max. = d$ *subject to the subsidiary conditions* $f_x + \lambda g_x = 0, f_y + \lambda g_y = 0$.

The problem $f = (x + 1)^2 + y^2 = \text{min.}$ subject to the condition $g = 2x = 0$ may serve as an illustration of our maximum-minimum

considerations. Geometrically the problem is to find the lowest point or vertex of the vertical parabola which is formed by the intersection of the paraboloid $z = (x + 1)^2 + y^2$ with the plane $x = 0$. We obtain immediately the value $d = 1$ for the desired least value of $z = (x + 1)^2 + y^2$. We now note that for fixed λ the paraboloid $z = f + \lambda g = (x + \lambda + 1)^2 + y^2 - 2\lambda - \lambda^2$ always contains the above parabola and that the vertex of the paraboloid lies lower than that of the parabola. By varying λ we vary the vertex of the paraboloid which will at most move up to the vertex of the parabola but never any higher. Thus the vertex of the parabola is the highest point of the set of lowest points of our paraboloids.

2. Involutory Transformation of the Simplest Variational Problems. Analogous transformations of variational problems are based upon the following general principle: *If a functional $J[u, v, \cdots]$ is stationary for a certain admissible system of functions $u, v, \cdots$, which may be required to fulfill certain subsidiary conditions, then J remains stationary for this system of functions when the set of subsidiary conditions is enlarged to include any further relations already satisfied by the functions $u, v, \cdots$.*

We shall call conditions that are necessary for the vanishing of the variation (such as the Euler equations and the natural boundary conditions) *natural conditions*; subsidiary and boundary conditions imposed a priori will be known as *constraints*. Then, from our principle: *If a variational problem for a given functional is changed by the explicit addition of one or more natural conditions to the set of constraints, the stationary character of the functional is not affected.*

We shall now turn to some problems of the simplest type:

I: $J = \int_{x_0}^{x_1} F(x, u, u')\,dx$ *is to have a stationary value subject to the usual continuity conditions, the boundary conditions*

$$u(x_0) - u_0 = 0, \qquad u(x_1) - u_1 = 0, \tag{80}$$

and the subsidiary condition

$$\frac{du}{dx} - u' = 0. \tag{81}$$

That is, we regard the variational problem as one involving two unknown functions u and u' subject to the differential equation of constraint (81). In accordance with the multiplier rule the solutions of

I are simultaneously solutions of the following problem

$$\text{II:} \quad \begin{aligned} &H[u, u', \lambda; \mu_0, \mu_1] \\ &= \int_{x_0}^{x_1} \left[F + \lambda\left(\frac{du}{dx} - u'\right)\right] dx - \mu_0[u(x_0) - u_0] + \mu_1[u(x_1) - u_1] \end{aligned}$$

is to be made stationary, where $u(x)$, $u'(x)$, $\lambda(x)$ *and the parameters* μ_0 *and* μ_1 *are to be determined, and there are no subsidiary or boundary conditions, i.e. the problem is free.* The variational equations, i.e. the Euler equations and the natural boundary conditions of the problem, are

$$F_{u'} - \lambda = 0, \tag{82}$$

$$F_u - \frac{d\lambda}{dx} = 0, \tag{83}$$

$$\frac{du}{dx} - u' = 0 \tag{84}$$

in the interior of the interval, and

$$\lambda(x_0) + \mu_0 = 0, \qquad \lambda(x_1) + \mu_1 = 0, \tag{85}$$

$$u(x_0) - u_0 = 0, \qquad u(x_1) - u_1 = 0 \tag{86}$$

at the end points, as is found immediately when the first variation is set equal to zero. If we eliminate λ, μ_0, μ_1, we obtain the Euler equation.

If we apply our general principle to problem II by appending to the set of constraints the conditions $du/dx - u' = 0$, $u(x_0) - u_0 = 0$, $u(x_1) - u_1 = 0$, we are led back to problem I. If, on the other hand, we adjoin equations (82), (83), (85), which correspond to the natural conditions of problem I, we obtain a transformation which is of importance in applications; we shall call it the *transformation into the reciprocal form of the variational problem.*[1] In this way we obtain problem III, which may be expressed as a problem of type I if we remove the derivative du/dx from the integral H by partial integration and then introduce new argument functions p, p' and a new integrand $\Psi(x, p, p')$ by means of equations

$$F_{u'} = p, \qquad F_u = p', \qquad pu' + p'u - F = \Psi. \tag{87}$$

[1] The significance of this transformation was first discovered by K. O. Friedrichs, Ein Verfahren der Variationsrechnung, . . . , Nachr. der Ges. d. Wiss., Göttingen, 1929, pp. 13–20.

In order that this "Legendre transformation" be meaningful we must require that u and u' can be obtained from the first two equations as functions of p, p', and x; these functions are then to be substituted in the left side of the third equation. This procedure certainly can be carried out, if the requirement

$$F_{u'u'}F_{uu} - (F_{uu'})^2 \neq 0 \tag{88}$$

is fulfilled for all sets of values x, u, u' of the fundamental domain. We thus obtain the following "reciprocal" problem,[1] equivalent to I,

$$\text{IV:} \qquad -\int_{x_0}^{x_1} \Psi(x, p, p')\, dx + p(x_1)u_1 - p(x_0)u_0$$

is to be made stationary under the subsidiary condition

$$\frac{dp}{dx} - p' = 0;$$

no boundary conditions are imposed.

The natural conditions for problem IV are

$$\frac{d}{dx}\Psi_{p'} - \Psi_p = 0$$

in the interior and

$$\Psi_{p'}|_{x_0} - u_0 = 0, \qquad \Psi_{p'}|_{x_1} - u_1 = 0$$

on the boundary. It is apparent from their derivation that these constraints are identical with those of problem I; this may be independently verified with the help of the inversion

$$\Psi_{p'} = u, \qquad \Psi_p = u', \qquad up' + u'p - \Psi = F$$

of the Legendre transformation (87).

By use of the same formulas we see that the reciprocal transformation applied to the free problem IV leads back to the original problem I. *Thus this transformation is of involutory character, and the natural conditions of the one problem go over into the constraints of the other.*

The degenerate case in which the integrand F of the variational problem does not depend on u explicitly or contains u only linearly requires special treatment. We have here

$$F(x, u, u') = g(x, u') + uf(x).$$

[1] This may be regarded as the analogue of problem IV of subsection 1.

In this case the Legendre transformation considered previously is not reversible for arbitrary p and p'. But by applying our transformation principle directly we find that the following variational problem is equivalent to the original problem I:

$$-\int_{x_0}^{x_1} \Phi(p)\,dx + p(x_1)u_1 - p(x_0)u_0 = \text{stationary},$$

with the subsidiary condition $dp/dx = f(x)$. Here p and $\Phi(p)$ are related to the expressions in the original problem by the transformation

$$p = g_{u'}, \qquad -\Phi(p) = g(x, u') - u'p,$$

and it is assumed that the equation $g_{u'} = p$ can be solved for u' in terms of p. The new problem is essentially simpler than the original one in the following respect: the required function $p(x)$ can be obtained, aside from an additive parameter, by quadrature from the subsidiary condition. Thus in this degenerate case *the variational problem goes over into an ordinary extremum problem* of determining a parameter.

As in subsection 1 we now examine the effect of these transformations upon the maximum or minimum character of our expressions.

By repeating the arguments of subsection 1, we shall arrive at the following result: *If the original problem I (henceforth to be denoted by I') pertains to a minimum d, then the same value d will occur as a maximum in the corresponding reciprocal problem IV (to be denoted by IV').*

Again, this statement is true only if certain restrictions are imposed. We require, specifically, that for any arbitrary $\lambda(x)$ with a piecewise continuous derivative, for which $\lambda(x_1) + \mu_1 = 0$, $\lambda(x_0) + \mu_0 = 0$,[1] the expression H of problem II, page 234, possess a minimum d_λ depending on λ. Then, removing the derivative of u by partial integration, we obtain problem

$$H = \int_{x_0}^{x_1}\left[F + \lambda\left(\frac{du}{dx} - u'\right)\right]dx$$

II′:
$$- \mu_0(u(x_0) - u_0) + \mu_1(u(x_1) - u_1)$$

$$= \int_{x_0}^{x_1}\left[F - \frac{d\lambda}{dx}\,u - \lambda u'\right]dx - \lambda(x_0)u_0 + \lambda(x_1)u_1$$

[1] Without this condition, which is automatically satisfied by the solution, it would certainly be impossible to obtain a minimum with arbitrary λ, μ.

is to be made a minimum, $\lambda(x)$ being a fixed function. The functions $u(x)$, $u'(x)$ that solve this problem then satisfy equations

$$F_{u'} - \lambda = 0, \qquad F_u - \frac{d\lambda}{dx} = 0. \tag{89}$$

We now assume, in analogy with the procedure of subsection 1, that these equations determine the functions u and u' uniquely for arbitrary values of λ and $d\lambda/dx$.

Since problem I′ arises from problem II′ when we impose the additional subsidiary conditions $du/dx - u' = 0$, $u(x_0) - u_0 = 0$, $u(x_1) - u_1 = 0$, we certainly have $d \geq d_\lambda$.

On the other hand equations (89) are satisfied by the solution of problem I′ with $\lambda = \bar{\lambda} = F_{u'}$ and, because of the assumption of uniqueness, we have $d_{\bar{\lambda}} = d$.

It follows that

$$d = \max d_\lambda .$$

But the problem of maximizing d_λ is just the problem IV′, and thus our assertion is proved.

A sufficient criterion for the validity of our assumptions is that the inequalities

$$F_{u'u'}F_{uu} - (F_{uu'})^2 > 0, \qquad F_{u'u'} > 0 \tag{90}$$

hold for all u and x in the domain under discussion and for arbitrary u'. As we have already seen (page 216), if these inequalities hold a solution of the Euler equations furnishes a minimum for problem I′. The existence of a minimum d_λ in problem II′ follows from these inequalities similarly; for, equations (89) and inequalities (90) together express the fact that at each value of x the integrand of H possesses a minimum for the appropriate values of u and u'. Certainly, then, H itself is a minimum.

In conclusion we may point out that the transition from a minimum problem to a maximum problem, as accomplished by the reciprocal transformation, may be exhibited directly under assumption (90). In the following reasoning, the reciprocal transformation will be obtained once more. From Taylor's expansion, in view of the inequalities (90), we obtain at once the inequality

$$F(u, u') - F(v, v') - (u - v)F_v - (u' - v')F_{v'} \geq 0,$$

in which the equality holds if and only if $u = v$, $u' = v'$. If we write the above expression in the form

$$F(u, u') - [F(v, v') - vF_v - v'F_{v'}] - uF_v - u'F_{v'}$$

and replace v and v' by the quantities p and p' by means of the Legendre transformation

$$p = F_{v'}, \qquad p' = F_v, \qquad \Psi(x, p, p') = vp' + v'p - F,$$

the inequality becomes

$$F(x, u, u') + \Psi(x, p, p') - up' - u'p \geq 0$$

for arbitrary u, u', p, p'; equality occurs if and only if p and p' correspond to the functions $v = u$, $v' = u'$. We now integrate this inequality between the limits x_0 and x_1, regarding u, u', p, p' as functions of x subject to the constraints

$$\frac{du}{dx} - u' = 0, \qquad \frac{dp}{dx} - p' = 0, \qquad \begin{matrix} u(x_0) - u_0 = 0 \\ u(x_1) - u_1 = 0. \end{matrix}$$

The left side can certainly not be negative; it vanishes if and only if u is the solution of problem I′ and p the solution of problem IV′. Thus the problem

$$\int_{x_0}^{x_1} [F + \Psi - up' - u'p]\, dx$$

$$= \int_{x_0}^{x_1} F\, dx + \int_{x_0}^{x_1} \Psi\, dx + u_0 p(x_0) - u_1 p(x_1) = \text{min.}$$

with the given constraints possesses this solution, and the minimum value in this problem is zero. This statement is equivalent to the above assertion concerning the relation of the two problems I′ and IV′.

3. Transformation of Variational Problems to Canonical Form. The general principle formulated in subsection 2 leads to another well-known transformation, the *transformation into canonical form.* Here the Euler second order differential equation is replaced by a system of differential equations of first order. This transformation, which has no exact counterpart in subsection 1, is obtained by imposing equations (82) and (86) as constraints in problem II. Thus one first obtains the problem

$$\text{IIa:} \qquad \int_{x_0}^{x_1} \left[F(x, u, u') + F_{u'} \left(\frac{du}{dx} - u' \right) \right] dx$$

is to be made stationary with the boundary conditions $u(x_0) = u_0$, $u(x_1) = u_1$, where u and u' are to be regarded as two independent argument functions.

If we introduce a new argument function[1]

$$p = F_{u'}$$

in place of u' and a new integrand

$$\Phi(x, u, p) = pu' - F(x, u, u')$$

in place of $F(x, u, u')$—we expressly assume

$$F_{u'u'} \neq 0,$$

so that u' can be determined as a function of p, u, and x from the relation $p = F_{u'}$—we obtain the equivalent problem

$$\text{IIb:} \qquad \int_{x_0}^{x_1} \left[p \frac{du}{dx} - \Phi(x, u, p) \right] dx = \text{stationary}$$

with the boundary conditions $u(x_0) = u_0$, $u(x_1) = u_1$. It is easily seen that the quantities occurring in the equivalent problems I and IIb are connected by the *Legendre transformation*

$$F_{u'} = p, \qquad pu' - F = \Phi,$$

whose inverse is expressed by the equations

$$\Phi_p = u', \qquad pu' - \Phi = F.$$

This form of the variational problem is known as the *canonical form*. The formulation of the variational equations for p and u yields the *canonical differential equations* of the variational problem:

$$\frac{dp}{dx} + \Phi_u = 0, \qquad \frac{du}{dx} - \Phi_p = 0.$$

In an analogous manner we may transform a variational problem for n unknown functions $u_1(x)$, $u_2(x)$, $\cdots$, $u_n(x)$ of the independent variable x into canonical form.

Summarizing the above discussion, we state without explicitly repeating qualifying sufficient conditions: Suppose, in problem I, d is a minimum; then, in the canonical problem, d becomes a maximum-minimum if for fixed p we find a minimum by varying u and then

[1] Thus p is equal to the multiplier occurring in II

determine the maximum of this minimum (a function of p) by varying p.

4. Generalizations. Our transformation theory is easily extended to problems involving several unknown functions, higher derivatives, or several independent variables. We confine ourselves here to the treatment of a particularly simple example which corresponds to the degenerate case of subsection 2, namely the transformation of Dirichlet's classical variational problem

$$\text{I:} \qquad \frac{1}{2}\iint_G \left[\left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial u}{\partial y}\right)^2\right] dx\,dy = \min.$$

in which u is a function of x, y with piecewise continuous derivatives in the region G and prescribed boundary values $\bar{u} = f(s)$. The boundary Γ of G is assumed to be a curve with continuously turning tangent (except possibly at a finite number of points) and with arc length s.

If in problem I we replace the two partial derivatives by the functions p and q and append the subsidiary conditions $\partial u/\partial x = p$, $\partial u/\partial y = q$, the multiplier rule immediately leads to the equivalent problem

$$\text{II:} \qquad \iint_G \left[\tfrac{1}{2}(p^2 + q^2) + \lambda\left(\frac{\partial u}{\partial x} - p\right) + \mu\left(\frac{\partial u}{\partial y} - q\right)\right] dx\,dy - \int_\Gamma \rho(s)[\bar{u} - f(s)]\,ds = \text{stat.}$$

Here $\lambda(x, y)$, $\mu(x, y)$, $\rho(s)$ are multipliers. Transforming the double integral by partial integration we obtain II in the form

$$\iint_G \left[\tfrac{1}{2}(p^2 + q^2) - u\left(\frac{\partial \lambda}{\partial x} + \frac{\partial \mu}{\partial y}\right) - \lambda p - \mu q\right] dx\,dy + \int_\Gamma \left[\bar{u}\left(\bar{\lambda}\frac{\partial x}{\partial n} + \bar{\mu}\frac{\partial y}{\partial n} - \rho(s)\right) + \rho(s) f(s)\right] ds = \text{stat.}$$

Here $\partial/\partial n$ denotes differentiation with respect to the outer normal, and boundary values are indicated by a bar ($\bar{u}$, $\bar{\lambda}$, $\bar{\mu}$). We now impose as explicit subsidiary conditions the Euler differential equations and certain natural boundary conditions:

$$p - \lambda = 0, \qquad q - \mu = 0,$$

$$\frac{\partial \lambda}{\partial x} + \frac{\partial \mu}{\partial y} = 0, \qquad \bar{\lambda}\frac{\partial x}{\partial n} + \bar{\mu}\frac{\partial y}{\partial n} - \rho = 0,$$

thus obtaining the equivalent problem

$$\text{III:} \qquad -\tfrac{1}{2}\iint_G (p^2 + q^2)\,dx\,dy + \int_\Gamma \rho(s) f(s)\,ds = \text{stat.}$$

with the subsidiary conditions

$$\rho(s) - \bar{p}\frac{\partial x}{\partial n} - \bar{q}\frac{\partial y}{\partial n} = 0$$

on the boundary and

$$\frac{\partial p}{\partial x} + \frac{\partial q}{\partial y} = 0$$

in the interior. The reciprocal transformation is completed.

We can simplify the result if we satisfy the latter subsidiary conditions by introducing a function $v(x, y)$ such that

$$p = \frac{\partial v}{\partial y}, \qquad q = -\frac{\partial v}{\partial x}.$$

Then

$$\bar{p}\frac{\partial x}{\partial n} + \bar{q}\frac{\partial y}{\partial n} = \frac{\partial v}{\partial s},$$

where the differentiation on the right is to be performed in the direction of the positive tangent of Γ; the problem goes over into problem

$$\text{IV:} \quad -\frac{1}{2}\iint_G \left[\left(\frac{\partial v}{\partial x}\right)^2 + \left(\frac{\partial v}{\partial y}\right)^2\right] dx\,dy + \int_\Gamma \frac{\partial v}{\partial s} f(s)\,ds = \text{stat.}$$

The boundary integral on the right, incidentally, may be transformed by partial integration into the form

$$-\int_\Gamma v f'(s)\,ds.$$

In this new problem, the integrand is of the same form as in problem I. The solution of problem I defines a function satisfying the potential equation $\Delta u = \partial^2 u/\partial x^2 + \partial^2 u/\partial y^2 = 0$ with the boundary values $f(s)$; problem IV also defines a potential function v which, in virtue of the natural boundary conditions, turns out to be the potential function conjugate to u.

A minimum in problem I corresponds to a maximum having the

same value in problem IV. This becomes clear if we subtract the expression IV from I; a simple transformation leads to the expression

$$\iint_G \left[\left(\frac{\partial u}{\partial x} - \frac{\partial v}{\partial y} \right)^2 + \left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x} \right)^2 \right] dx\, dy.$$

Problems I and IV together are equivalent to the problem of minimizing this last integral subject to the single boundary condition $\bar{u} = f(s)$. The minimum has the value zero and is attained when u is the solution of the appropriate boundary value problem of potential theory and v is its conjugate function satisfying the differential equations

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}, \qquad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}.$$

For a more direct and general treatment of reciprocal quadratic variational problems see §11.

§10. *Variational Calculus and the Differential Equations of Mathematical Physics*

1. General Remarks. The calculus of variations is a reliable guide both in formulating and in treating differential equations of mathematical physics. Problems of (stable) equilibrium are governed by the variational principle of the *minimum potential energy*; the laws of motion are most simply formulated in terms of *Hamilton's variational principle*. By these two principles we shall derive some of the fundamental differential equations of mathematical physics.

Let us first consider a system with a finite number, say n, of degrees of freedom. Let the position of the system be characterized by the values of n parameters $q_1, q_2, \cdots, q_n$; the object is to determine these as functions of the time t. We suppose that the mechanical properties of the system are determined by two quantities, the kinetic and potential energy. We assume that the *kinetic energy* $T(\dot{q}_1, \dot{q}_2, \cdots, \dot{q}_n, q_1, q_2, \cdots, q_n, t)$ is a known function of the n velocities $\dot{q}_i$, the n coordinates q_i, and the time t, specifically that it is a quadratic form in the velocities:

$$T = \sum_{i,k=1}^{n} P_{ik}(q_1, q_2, \cdots, q_n, t)\dot{q}_i\, \dot{q}_k.$$

The *potential energy* $U(q_1, q_2, \cdots, q_n, t)$ is assumed to be a known function of t and the coordinates q_i. Then Hamilton's principle

states: *Between two instants of time t_0 and t_1 the motion proceeds in such a way that for the functions $q_i(t)$ the integral*

$$J = \int_{t_0}^{t_1} (T - U)\, dt$$

is stationary in comparison with the neighboring functions $\bar{q}_i(t)$ for which $\bar{q}_i(t_0) = q_i(t_0)$ and $\bar{q}_i(t_1) = q_i(t_1)$. In other words: *The actual motion makes the value of the integral J stationary with respect to all neighboring virtual motions which lead from the initial to the final position of the system in the same interval of time.*

In accordance with §3 Hamilton's principle leads at once to *Lagrange's general equations of motion*

$$\frac{d}{dt}\frac{\partial T}{\partial \dot{q}_i} - \frac{\partial}{\partial q_i}(T - U) = 0 \qquad (i = 1, 2, \cdots, n). \tag{91}$$

The *equilibrium conditions* can be obtained from the equations of motion, under the assumption that T and U do not depend explicitly on t, by setting the derivatives with respect to time equal to zero in (91). The resulting conditions are

$$\frac{\partial U}{\partial q_i} = 0. \tag{92}$$

Thus: A mechanical system with the potential energy $U(q_1, q_2, \cdots, q_n)$ is in equilibrium at a particular set of values of the coordinates $q_1, q_2, \cdots, q_n$ if and only if the potential energy is stationary for this set of values.

In order that the equilibrium be stable, it is, moreover, necessary and sufficient that the stationary value of U in the equilibrium position be a minimum.

This fact follows easily from the principle of conservation of energy, $T + U = \text{const.}$, an immediate consequence of equations (91). We shall henceforth accept the principle of minimum potential energy as the condition of stable equilibrium.

A motion takes on a particularly simple character if confined to the vicinity of a stable equilibrium position of a system. We may assume without loss of generality that at the equilibrium in question all the coordinates q_i vanish. If we now consider those motions close to the equilibrium position for which higher powers of the coordinates and of their time derivatives are negligible compared with lower powers, and if we assume that T and U do not contain

t explicitly, then we may regard T as a positive definite quadratic form in $\dot{q}_i$ with constant coefficients a_{ik} :

$$T = \sum_{i,k=1}^{n} a_{ik}\dot{q}_i\dot{q}_k .$$

Under these hypotheses, U also is a positive definite quadratic form in q_i with the constant coefficients b_{ik} :

$$U = \sum_{i,k=1}^{n} b_{ik}q_iq_k .$$

Thus the equations of motion go over into the linear second order differential equations with constant coefficients

$$\sum_{k=1}^{n} a_{ik}\ddot{q}_k + \sum_{k=1}^{n} b_{ik}q_k = 0$$

which govern "*small vibrations*" about a stable equilibrium position, and which will be discussed in detail in the next chapter.

In the case of problems of continuum mechanics, in which the position of the system no longer can be described by a finite number of variables, we may also proceed from Hamilton's principle or from the principle of least potential energy. Here the kinetic and potential energies are not functions of a finite number of variables, but are functionals represented by integrals over appropriate space regions, surfaces, or lines.

2. The Vibrating String and the Vibrating Rod. The simplest example of continuum mechanics is given by the homogeneous vibrating string, subjected to a constant tension μ, which executes small transverse vibrations about the position of stable equilibrium, the interval $0 \leq x \leq l$ of the x-axis. Our problem is to determine the perpendicular deflection $u(x, t)$ of a point on the string from its equilibrium position. We assume that the motion is "small" in the sense, specifically, that higher powers of the function u and of its derivatives may be neglected compared with lower powers. First, let us consider the case in which the string is fixed at its end points, i.e. $u(0, t) = u(l, t) = 0$. The kinetic energy of the string is given by the integral $T = \frac{1}{2}\int_0^l \rho u_t^2\, dx$, where ρ denotes the linear density of the string. The potential energy U is proportional to the increase in length compared with the length at rest, the factor of propor-

tionality being equal to the tension μ. Since, apart from quantities of higher order, the change in length $\int_0^l \sqrt{1 + u_x^2}\, dx - l$ approximately equals $\frac{1}{2}\int_0^l u_x^2\, dx$, the potential energy takes the form

$$U = \tfrac{1}{2}\int_0^l \mu u_x^2\, dx.$$

Hamilton's principle, therefore, leads to the problem of finding an admissible function $u(x, t)$ for which the double integral

$$\int_{t_0}^{t_1} (T - U)\, dt = \tfrac{1}{2}\int_{t_0}^{t_1}\int_0^l (\rho u_t^2 - \mu u_x^2)\, dx\, dt$$

is stationary compared with all the continuous functions $u(x, t)$ which have piecewise continuous first derivatives, vanish at $x = 0$ and at $x = l$, and coincide with the functions $u(x, t_0)$ and $u(x, t_1)$ corresponding to the actual motion at $t = t_0$ and $t = t_1$, respectively. For constant ρ and μ, we then obtain from the general laws of the variational calculus the partial differential equation of the vibrating string

$$\rho u_{tt} - \mu u_{xx} = 0. \tag{93}$$

If an external force $f(x, t)$ acts on the string the term $\int_0^l f(x, t)u\, dx$ must be added to the potential energy, and we now obtain the differential equation

$$\rho u_{tt} - \mu u_{xx} + f(x, t) = 0. \tag{94}$$

The stable equilibrium of a string under the influence of an external force is given, in accordance with the minimum principle, by the minimum of the integral $\int_0^l (\frac{1}{2}\mu u_x^2 + fu)\, dx$, where the external force $f(x)$ is, of course, assumed to be time-independent. This leads to the Euler equation

$$\mu u_{xx} - f(x) = 0$$

as a special case of the equation of motion (94).

We now formulate the corresponding equations for the states of a laterally movable rod. The rod is defined as a one-dimensional continuum, lying in a straight line when at rest, whose potential energy

at deformation is proportional to the integral of the square of the curvature, extended over its length. If we again assume that higher powers of the deformation function $u(x, t)$ and of its derivatives are negligible compared with lower powers, we obtain an expression of the form $\frac{1}{2}\mu \int_0^l u_{xx}^2\, dx$ for the potential energy of deformation. The kinetic energy has the same form as that of the string. With an external force $f(x, t)$ acting, the equation of motion, derived from Hamilton's principle, becomes

$$\rho u_{tt} + \mu u_{xxxx} + f(x, t) = 0,$$

and the condition for equilibrium under an external force $f(x)$ is

$$\mu u_{xxxx} + f(x) = 0.$$

The boundary conditions or other constraints to be imposed are of great importance in the solution of our variational problems. We may, for example, fix the boundary by conditions such as $u(0) = u(l) = 0$ for the string or $u(0) = u_x(0) = u(l) = u_x(l) = 0$ for the rod, or we may instead leave the boundaries free. For free boundaries, the methods of §5 lead to the natural boundary conditions

$$u_x(0, t) = u_x(l, t) = 0 \tag{95}$$

for the string and

$$u_{xx}(0, t) = u_{xx}(l, t) = 0, \qquad u_{xxx}(0, t) = u_{xxx}(l, t) = 0 \tag{96}$$

for the rod.

If the end points of the string are neither fixed nor free but are held by elastic forces, boundary terms $\frac{1}{2}h_1\mu u^2(0, t)$ and $\frac{1}{2}h_2\mu u^2(l, t)$ must be added to the potential energy. These terms do not alter the equation of motion (94), but lead to the natural boundary conditions[1]

$$u_x(0, t) = h_1 u(0, t), \qquad u_x(l, t) = -h_2 u(l, t).$$

3. Membrane and Plate. For the plane membrane and plate the situation is similar to that for the string and rod. A membrane is a portion of a surface, plane at rest, with potential energy proportional to change in area; the proportionality factor is known

[1] For further remarks concerning rods, see §12, 12–13.

as the tension. Suppose that the membrane at rest covers a region G of the x, y-plane; let the deformation normal to the equilibrium plane be denoted by $u(x, y)$, and suppose this deformation to be small in the sense that higher powers of u, u_x, u_y are negligible compared with lower ones. Then the expression $\iint_G (1 + u_x^2 + u_y^2)^{\frac{1}{2}}\, dx\, dy$ for the area may be replaced by $\iint_G [1 + \frac{1}{2}(u_x^2 + u_y^2)]\, dx\, dy$, and the required potential energy, apart from a constant factor, is given by the double integral

$$\frac{1}{2} \iint_{\blacksquare} (u_x^2 + u_y^2)\, dx\, dy. \tag{97}$$

We consider first the *equilibrium problem for the membrane.* If we suppose that the displacement $u(x, y)$ of the membrane possesses prescribed values $\bar{u} = \bar{u}(s)$ on the boundary Γ of G—where s denotes the arc length of Γ—and that no external forces act on the membrane, then the equilibrium position is characterized by the following variational problem: The displacement $u(x, y)$ in the equilibrium position is that function for which the integral $\iint_{\blacksquare} (u_x^2 + u_y^2)\, dx\, dy$ attains the least possible value if all those functions u are admitted for competition which are continuous in the closed domain G, take on the prescribed boundary values $\bar{u}(s)$, and possess continuous first and piecewise continuous second derivatives[1] in the interior. The Euler equation is

$$\Delta u = u_{xx} + u_{yy} = 0.$$

Thus the problem of finding the equilibrium position is equivalent to the *boundary value problem*: find in G a solution of the above *partial differential equation* (the *potential equation*) which assumes prescribed values on the boundary Γ of G.

We now make the more general assumptions that the interior of the membrane is subject to an external force of surface density $f(x, y)$, the boundary of the membrane (supposed to be freely movable above Γ) to an external force of linear density $p(s)$, and, finally,

[1] Later, for the complete treatment of the problem, it will be important that the continuity requirements for the second derivatives may be relaxed without altering the solution.

that the boundary is tied to its rest position by elastic forces which are characterized by a modulus of elasticity of linear density $\sigma(s)$. Then the potential energy of a membrane with the displacement $u(x, y)$ is given by the expression

$$\iint_G [\tfrac{1}{2}\mu(u_x^2 + u_y^2) + fu]\, dx\, dy + \int_\Gamma [p(s)u + \tfrac{1}{2}\sigma(s)u^2]\, ds.$$

Once more we obtain the equilibrium position by finding a function $u(x, y)$, subject to no boundary conditions but only to the above continuity conditions, for which this integral is a minimum. The Euler equation (which expresses the equilibrium condition in the interior) is, if we take $\mu = 1$,

$$\Delta u = f(x, y),$$

and the natural boundary condition is

$$\frac{\partial u}{\partial n} + \sigma u + p(s) = 0.$$

These two requirements again represent a boundary value problem for a partial differential equation.

From this general case, we again arrive at the problem of solving the differential equation $\Delta u = f$ with the boundary values $u = 0$, if we set p equal to zero and let σ increase beyond all bounds.

If $\sigma = 0$ our equilibrium problem does not in general possess a solution. Physically, this is plausible; for, a freely movable membrane, subjected to arbitrary external forces, cannot have a stable equilibrium position unless the external forces happen to be balanced. This fact is easily proved from the variational principle: In order that our expression for the energy possess a lower bound in the case $\sigma = 0$, equation

$$\iint_G f\, dx\, dy + \int_\Gamma p\, ds = 0 \tag{98}$$

must be satisfied. Indeed, for a constant displacement $u(x, y) = c$, the energy is c times the left side of (98) and can therefore have arbitrarily large negative values, unless the left side vanishes. But if we impose condition (98), the solution of the variational or equilibrium problem is not determined uniquely, since an arbitrary constant can be added to u without altering the value of the energy; its

minimum is therefore also unchanged. We must impose a further condition in order to fix the solution uniquely. It is customary to use the condition

$$\iint_B u \, dx \, dy = 0,$$

which means that the center of mass of the membrane is fixed in the rest position.

We arrive at the equations of motion of a membrane by Hamilton's principle, noting that the kinetic energy is given by

$$T = \tfrac{1}{2}\rho \iint_B u_t^2 \, dx \, dy. \tag{99}$$

If external forces act with surface density $f(x, y, t)$ in the interior of the membrane and line density $p(s, t)$ on the boundary, and if the boundary is again subjected to elastic forces with the elastic modulus $\sigma(s)$, Hamilton's principle requires that the expression

$$\int_{t_0}^{t_1} \iint_G [\tfrac{1}{2}\rho u_t^2 - \tfrac{1}{2}\mu(u_x^2 + u_y^2) - f(x, y, t)u] \, dx \, dy \, dt$$

$$- \int_{t_0}^{t_1} \int_\Gamma (\tfrac{1}{2}\sigma u^2 + pu) \, ds \, dt$$

be made stationary. The Euler equation for this problem is

$$\mu \Delta u - \rho u_{tt} - f(x, y, t) = 0;$$

the natural boundary conditions are

$$\frac{\partial u}{\partial n} + \sigma u + p(s, t) = 0. \tag{100}$$

If the membrane has a fixed boundary, i.e. if the values of u on the boundary are given as a function of the arc length, then the prescription of these boundary values takes the place of condition (100).

In the equilibrium case the minimum principle not only gives the pertinent formulation but also proves a powerful tool for solving and analyzing the problem. However, in dynamic problems the chief use of Hamilton's principle is to derive formally the differential equation; for the further analysis of these equations Hamilton's principle is often not suitable since it requires that the functions ad-

mitted for comparison be prescribed at two fixed instants of time t_0 and t_1, conditions not generally encountered in actual problems. The information usually given in dynamical problems comprises boundary conditions and, in addition, *initial conditions*, i.e. the values of the functions $u(x, y, t)$ and $u_t(x, y, t)$ at *one* instant $t = 0$. Thus dynamic problems lead to mixed boundary and initial value problems.

The situation is analogous in the case of the plate. A plate is an elastic two-dimensional body, plane when in equilibrium, whose potential energy under deformation is given by an integral of a quadratic form in the principal curvatures of the plate. If the principal radii of curvature of the deformed plate are denoted by ρ_1, ρ_2 the potential energy density is given by an expression of the form

$$A\left(\frac{1}{\rho_1^2} + \frac{1}{\rho_2^2}\right) + \frac{2B}{\rho_1\rho_2},$$

where A and B are constants determined by the material of the plate. Assuming u, u_x, u_y, $\cdots$ to be small, we may write

$$\frac{2}{\rho_1} + \frac{2}{\rho_2} = \Delta u = u_{xx} + u_{yy}, \qquad \frac{1}{\rho_1\rho_2} = u_{xx}u_{yy} - u_{xy}^2.$$

Thus the desired potential energy of deformation is given by an expression of the form

$$U_1 = \iint_G [(\Delta u)^2 - 2(1 - \mu)(u_{xx}u_{yy} - u_{xy}^2)]\, dx\, dy, \tag{101}$$

apart from a constant factor (depending on the material) which we may set equal to unity.

To this must be added the energy due to external surface forces, boundary forces, and bending moments on the boundary if they are prescribed:

$$U_2 = \iint_G fu\, dx\, dy + \int_\Gamma p(s)u\, ds + \int_\Gamma m(s)\frac{\partial u}{\partial n}\, ds;$$

here $f(x, y)$, $p(s)$, $m(s)$ represent, respectively, the densities of applied surface forces, boundary forces, and bending moments normal to the boundary curve.

Equilibrium is once more characterized by the requirement that

$U_1 + U_2$ be a minimum for a suitable admissible function $u(x, y)$. (Here the admissible functions have continuous derivatives up to the fourth order, but these conditions may actually be greatly relaxed without affecting the solution of the problem.) To find the Euler equations and the natural boundary conditions for our minimum problem we must form the variation $\delta U = \delta U_1 + \delta U_2$, as outlined in §5, and equate it to zero. We obtain

$$\delta U_1 = \iint_G (\Delta\Delta u \, \delta u)\, dx\, dy - \int_\Gamma M \, \delta \frac{\partial u}{\partial n}\, ds - \int_\Gamma P \, \delta u \, ds,$$

with

$$M(u) = -\Delta u + (1 - \mu)(u_{xx}x_s^2 + 2u_{xy}x_s y_s + u_{yy}y_s^2),$$

$$P(u) = \frac{\partial}{\partial n}\Delta u + (1 - \mu)\frac{\partial}{\partial s}\left(u_{xx}x_n x_s + u_{xy}(x_n y_s + x_s y_n) + u_{yy}y_n y_s\right).$$

Here x_n, y_n and x_s, y_s are the direction cosines of the outward normal and the tangent vector, respectively. From the condition $\delta U = 0$ we obtain as the equilibrium conditions, the Euler differential equation

$$\Delta\Delta u + f = 0,$$

and, since no conditions have been imposed a priori on the boundary, the natural boundary conditions[1]

$$P(u) - p = 0, \qquad M(u) - m = 0.$$

If the plate is clamped at the edge, i.e. if u and $\partial u/\partial n$ have the prescribed value zero at the boundary, these natural boundary conditions are replaced by the conditions $u = 0$, $\partial u/\partial n = 0$. If the boundary of the plate is supported, i.e. if the boundary is fixed at zero while the tangent plane at the boundary is not constrained, the boundary conditions are

$$(102) \quad u = 0, \qquad \Delta u - (1 - \mu)(x_s^2 u_{xx} + y_s^2 u_{yy} + 2x_s y_s u_{xy}) = 0.$$

To formulate the differential equation for the motion of the plate (the differential equation of the vibrating plate) we again utilize

[1] It is worth noting that in the variational problem for the plate the expression $u_{xx}u_{yy} - u_{xy}^2$, because it is a divergence expression, does not affect the Euler differential equation; yet it is decisive for the form of the natural boundary conditions.

Hamilton's principle, employing the expression (99) for the kinetic energy. We obtain the equations

$$\mu\Delta\Delta u + \rho u_{tt} = 0$$

or, more generally,

$$\mu\Delta\Delta u + \rho u_{tt} + f(x, y, t) = 0$$

with appropriate boundary conditions just as above in the equilibrium problems. For the characterization of an actual physical problem these boundary conditions have to be supplemented by initial conditions which describe the initial state, i.e. the functions $u(x, y, 0)$ and $u_t(x, y, 0)$.

§11. *Reciprocal Quadratic Variational Problems*[1]

Linear functional equations of mathematical physics result from quadratic variational problems. It is illuminating, especially in regard to the reciprocity phenomenon discussed in §9, to consider such quadratic problems from a somewhat general and abstract viewpoint (compare Ch. I, §5, 3) and at the same time to use suggestive geometrical language.[2]

We consider, in a linear vector space Λ of vectors $\mathfrak{p}, \mathfrak{q}, \cdots$, a quadratic form $Q(\mathfrak{p}, \mathfrak{p})$ which we may assume to be positive definite. The vectors may be functions or systems of functions; for example

1) $$Q(\mathfrak{p}, \mathfrak{p}) = \iint_G \mathfrak{p}^2 \, dx \, dy$$

[1] This section has been added for the present edition.

[2] Recently several authors have rediscovered and advanced the theories indicated in §9. See J. L. Synge, The method of the hypercircle in function-space for boundary value problems, Proc. Roy. Soc. London, Ser. A, Vol. 191, 1941, pp. 447–467; W. Prager and J. L. Synge, Approximations in elasticity based on the concept of function space, Q. App. Math., Vol. 5, 1947, pp. 241–269; J. L. Synge, The method of the hypercircle in elasticity when body forces are present, Q. App. Math., Vol. 6, 1948, pp. 15–19 and the literature quoted there. Synge's papers have called attention to the advantage of a geometrical interpretation from which the distance of the approximations to the exact solutions can be deduced. See also J. B. Diaz and H. J. Greenberg, Upper and lower bounds for the solution of the first boundary value problem of elasticity, Q. App. Math., Vol. 6, 1948, pp. 326–331; H. J. Greenberg, The determination of upper and lower bounds for the solution of the Dirichlet problem, J. Math. Phys., Vol. 27, 1948, pp. 161–182; J. B. Diaz and H. J. Greenberg, Upper and lower bounds for the solution of the first biharmonic boundary value problem, J. Math. Phys., Vol. 27, 1948, pp. 193–201; J. B. Diaz, Upper and lower bounds for quadratic functionals, Proc. of the Symposium on Spectral Theory and Differential Problems, Stillwater, 1951.

where $\mathfrak{p} = \operatorname{grad} \varphi(x, y)$ in a region G, or

2) $$Q(\mathfrak{p}, \mathfrak{p}) = \iint_G \mathfrak{p}^2 \, dx \, dy$$

where $\mathfrak{p} = \Delta\varphi(x, y)$, or a form such as

3) $$Q(\mathfrak{p}, \mathfrak{p}) = \int_0^1 \int_0^1 K(x, y) p(x) p(y) \, dx \, dy \, .$$

The definition of the corresponding polar forms $Q(\mathfrak{p}, \mathfrak{q})$ is obvious.

We interpret $Q(\mathfrak{p}, \mathfrak{p})$ as the square of the length of $\mathfrak{p}$, and call vectors $\mathfrak{p}$, $\mathfrak{q}$ orthogonal if

$$Q(\mathfrak{p}, \mathfrak{q}) = 0.$$

For orthogonal vectors $\mathfrak{p}$, $\mathfrak{q}$ we have

$$Q(\mathfrak{p} + \mathfrak{q}, \mathfrak{p} + \mathfrak{q}) = Q(\mathfrak{p} - \mathfrak{q}, \mathfrak{p} - \mathfrak{q}) = Q(\mathfrak{p}, \mathfrak{p}) + Q(\mathfrak{q}, \mathfrak{q})$$

("Pythagorean Theorem").

If Ω is a linear subspace of the vector space Λ, we can define another linear subspace Σ orthogonal to Ω as the space of all vectors $\mathfrak{o}$ orthogonal to all vectors ω of Ω:

$$Q(\omega, \mathfrak{o}) = 0.$$

In all instances considered here the following facts are true: if Σ is the linear subspace orthogonal to Ω, then Ω is the linear subspace orthogonal to Σ. Every element $\mathfrak{p}$ in Λ can be uniquely decomposed into the sum of "projections" ω and $\mathfrak{o}$:

$$\mathfrak{p} = \omega + \mathfrak{o} \qquad (\omega \text{ in } \Omega, \ \mathfrak{o} \text{ in } \Sigma).$$

If we restrict our vectors by imposing suitable continuity and differentiability conditions, then the theorems which guarantee the existence of a solution for our variational problems also insure the validity of our statements. Accordingly, although we then operate in an incomplete Hilbert space, we assume that the projection theorem holds for the respective incomplete spaces.

We now consider two variational problems.

I: Given two vectors $\mathfrak{p}_0$ and $\mathfrak{q}_0$ in Λ, minimize

$$Q(\mathfrak{p}, \mathfrak{p}) - 2Q(\mathfrak{p}, \mathfrak{q}_0)$$

admitting for competition vectors $\mathfrak{p}$ for which $\mathfrak{p} - \mathfrak{p}_0$ is in a prescribed linear subspace Ω. Note that the difference of two admissible

vectors belongs to Ω, whereas $\mathbf{p}$ itself is restricted to a linear set Ω_0 of vectors "parallel" to the linear subspace Ω.

As a first example, let the vector space Λ consist of vectors $\mathbf{p}$ which are pairs of functions defined in a region G; the norm is given by

$$Q(\mathbf{p}, \mathbf{p}) = \iint_G \mathbf{p}^2 \, dx \, dy$$

Take $\mathbf{p}_0 = \text{grad } \varphi_0$ where φ_0 is a given function, and take as Ω the subspace of vectors $\boldsymbol{\omega} = \text{grad } \varphi$ with $\varphi = 0$ on the boundary Γ of G. The subspace Σ, the orthogonal complement of Ω, consists of those vectors $\boldsymbol{\sigma}$ for which $\text{div } \boldsymbol{\sigma} = 0$. If we choose $\mathbf{q}_0 = 0$, then for this example problem I is simply Dirichlet's problem: minimize

$\iint_G (\varphi_x^2 + \varphi_y^2) \, dx \, dy$ subject to the condition $\varphi - \varphi_0 = 0$ on Γ.

Next, let the vector space Λ consist of functions p defined in a region G; the norm is given by

$$Q(p, p) = \iint_G p^2 \, dx \, dy.$$

Take $p_0 = \Delta\varphi_0$ where φ_0 is a given function, and take as Ω the subspace of functions $\omega = \Delta\varphi$ with φ and its normal derivative vanishing on Γ. The subspace Σ then consists of those functions σ for which $\Delta\sigma = 0$. If we choose $q_0 = 0$, then we have for φ the variational problem associated with the clamped plate.

Suppose that $\mathbf{u}$ minimizes the functional of problem I. Since $\mathbf{p} = \mathbf{u} + \epsilon\boldsymbol{\omega}$ (for any number ϵ and any element $\boldsymbol{\omega}$ in Ω) is also admissible, the vanishing of the first variation amounts to the condition

$$Q(\mathbf{u} - \mathbf{q}_0, \boldsymbol{\omega}) = 0 \qquad \text{for all } \boldsymbol{\omega} \text{ in } \Omega.$$

Hence the equivalent of Euler's equation and the accompanying natural boundary conditions for problem I is that $\mathbf{u} - \mathbf{q}_0$ is an element of Σ, the orthogonal complement of Ω.

The uniqueness of the minimizing vector $\mathbf{u}$ is now apparent. For, if $\mathbf{u}'$ also furnishes a minimum, then, as we have just seen, $\mathbf{u}' - \mathbf{q}_0$ is in Σ; consequently $(\mathbf{u} - \mathbf{q}_0) - (\mathbf{u}' - \mathbf{q}_0) = \mathbf{u} - \mathbf{u}'$ is a vector of Σ. But, in view of the admissibility conditions, $\mathbf{u}$ and $\mathbf{u}'$ are in Ω_0 so that their difference $\mathbf{u} - \mathbf{u}'$ lies in Ω. Since the subspaces Ω and Σ contain only the vector zero in their intersection, we see that $\mathbf{u}' = \mathbf{u}$.

The reciprocal problem to problem I will now be formulated.

II: Minimize the expression

$$Q(\mathbf{q}, \mathbf{q}) - 2Q(\mathbf{q}, \mathbf{p}_0)$$

admitting for competition all vectors $\mathbf{q}$ for which $\mathbf{q} - \mathbf{q}_0$ is in Σ. If the minimum is attained for $\mathbf{q} = \mathbf{v}$, then by the same reasoning as above we can conclude that $\mathbf{v} - \mathbf{p}_0$ is in Ω.

Since $\mathbf{v} - \mathbf{q}_0$ is in Σ and $\mathbf{u} - \mathbf{q}_0$ is in Σ, the difference $\mathbf{u} - \mathbf{v}$ is in Σ; in the same way, from problem I we see that $\mathbf{v} - \mathbf{u}$ is in Ω. But since the two orthogonal subspaces have only the vector zero in common, it follows that $\mathbf{u} = \mathbf{v}$.

Thus problems I and II have the same solutions $\mathbf{u} = \mathbf{v}$.

By adding to the variational expressions in problems I and II the constant terms $Q(\mathbf{q}_0, \mathbf{q}_0)$ and $Q(\mathbf{p}_0, \mathbf{p}_0)$, respectively, we can state the problems in the following form:

I: Find the shortest distance from a fixed vector $\mathbf{q}_0$ to the linear set Ω_0, i.e. minimize

$$d(\mathbf{p}) = Q(\mathbf{p} - \mathbf{q}_0, \mathbf{p} - \mathbf{q}_0)$$

over all $\mathbf{p}$ in Ω_0.

II: Find the shortest distance from a fixed vector $\mathbf{p}_0$ to the linear set Σ_0, i.e. minimize

$$d(\mathbf{q}) = Q(\mathbf{q} - \mathbf{p}_0, \mathbf{q} - \mathbf{p}_0)$$

over all $\mathbf{q}$ in Σ_0.

Both minima d_1 and d_2 are attained by the same solution $\mathbf{p} = \mathbf{q} = \mathbf{u}$.

The reciprocal character of the two problems is expressed by the fact that the admissibility conditions of the one are the Euler conditions of the other.

Geometrically, the functions $\mathbf{p}$ and $\mathbf{q}$ are represented in Figure 2.

A glance at this figure and the theorems of Pythagoras or Thales suggest, moreover, the relations

$$d_1 + d_2 = Q(\mathbf{p}_0 - \mathbf{q}_0, \mathbf{p}_0 - \mathbf{q}_0)$$

and

$$Q(\mathbf{u} - \mathbf{p}, \mathbf{u} - \mathbf{p}) + Q(\mathbf{u} - \mathbf{q}, \mathbf{u} - \mathbf{q}) = 4Q\left(\mathbf{u} - \frac{\mathbf{p} + \mathbf{q}}{2}, \mathbf{u} - \frac{\mathbf{p} + \mathbf{q}}{2}\right)$$

where $\mathfrak{p}$, $\mathfrak{q}$ are any two vectors admissible respectively in the two reciprocal problems. These relations follow analytically from the fact that $\mathfrak{u} - \mathfrak{p}$ is in Ω, $\mathfrak{u} - \mathfrak{q}$ in Σ; hence $\mathfrak{u} - \mathfrak{p}$ and $\mathfrak{u} - \mathfrak{q}$ are orthogonal and

$$\begin{aligned} 4Q\left(\mathfrak{u} - \frac{\mathfrak{p}+\mathfrak{q}}{2}, \quad \mathfrak{u} - \frac{\mathfrak{p}+\mathfrak{q}}{2}\right) \\ &= Q(\mathfrak{u} - \mathfrak{p} + \mathfrak{u} - \mathfrak{q}, \mathfrak{u} - \mathfrak{p} + \mathfrak{u} - \mathfrak{q}) \\ &= Q(\mathfrak{u} - \mathfrak{p}, \mathfrak{u} - \mathfrak{p}) + Q(\mathfrak{u} - \mathfrak{q}, \mathfrak{u} - \mathfrak{q}) \\ &= Q(\mathfrak{p} - \mathfrak{q}, \mathfrak{p} - \mathfrak{q}). \end{aligned}$$

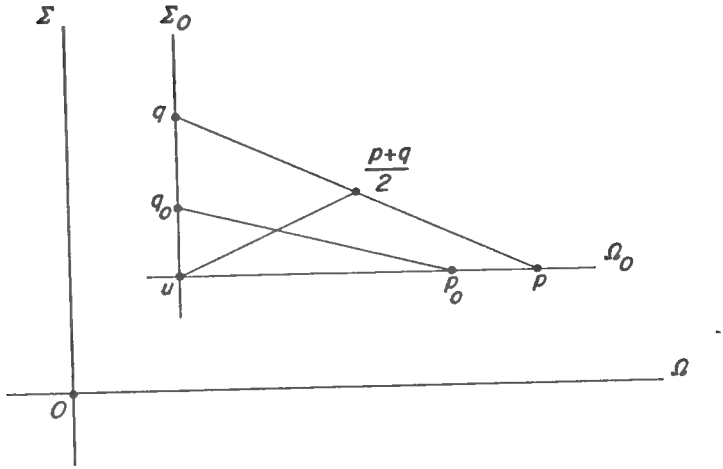


Figure 2

These considerations imply a remarkable fact: an upper bound for one of the two variational problems furnishes at the same time a lower bound for the reciprocal problem, since the sum of the minima is a priori known. This remark will prove relevant in connection with direct methods of the variational calculus.

Furthermore, the second relation,[1] corresponding to the theorem of Thales, states: The solution $\mathfrak{u}$ common to both problems deviates from the arithmetical mean of two arbitrary functions $\mathfrak{p}$, $\mathfrak{q}$ admissible, respectively, in the two reciprocal problems by an amount equal to half the distance between the two functions $\mathfrak{p}$ and $\mathfrak{q}$. Hence, once

[1] In the present connection this relation was pointed out by Synge and Prager.

$\mathbf{p}$ and $\mathbf{q}$ are chosen, the sphere of radius $|\ (\mathbf{p} - \mathbf{q})\ 2\ |$ about the point $(\mathbf{p} + \mathbf{q})/2$ is a geometric locus for $\mathbf{u}$.

Special cases (see also §12, 12) are easily fitted into the general scheme. For example, as we have seen, Dirichlet's problem for the harmonic differential equation in a region G with boundary Γ (see §8) corresponds to $Q(\mathbf{p}, \mathbf{p}) = \iint_G \mathbf{p}^2\, dx\, dy$. The space Ω_0 of $\mathbf{p}$ is defined by $\mathbf{p} = \text{grad}\ \varphi(x, y)$, $\varphi - \varphi_0 = 0$ on Γ, $\mathbf{p}_0 = \text{grad}\ \varphi_0(x, y)$, where φ_0 is a prescribed function. The space $\Sigma_0 = \Sigma$ of $\mathbf{q}$ is defined by $\text{div}\ \mathbf{q} = 0$, $\mathbf{q}_0 = 0$. Incidentally, it should be noted that in problem II the expression $Q(\mathbf{q}, \mathbf{p}_0)$ can be transformed into the boundary integral $\int_\Gamma \varphi_0 q_n\, ds$ where q_n is the normal component of $\mathbf{q}$.

Thus the reciprocal problem II can be formulated if we merely know the prescribed boundary values of φ; no explicit knowledge of a function $\varphi_0(x, y)$ is necessary.

A similar fact is true of other examples, e.g. the problem of the clamped plate, where

$$Q(\mathbf{p}, \mathbf{p}) = \iint_G \mathbf{p}^2\, dx\, dy,$$

$$\mathbf{p} = \Delta\varphi, \qquad \mathbf{p}_0 = \Delta\varphi_0, \qquad \Delta\mathbf{q} = 0, \qquad \mathbf{q}_0 = 0,$$

and where $\varphi - \varphi_0$ and its normal derivative are supposed to vanish on Γ. Since

$$Q(\mathbf{q}, \mathbf{p}_0) = \int_\Gamma \left(\varphi_0 \frac{\partial \mathbf{q}}{\partial n} - \frac{\partial \varphi_0}{\partial n}\, \mathbf{q} \right) ds,$$

where $\partial/\partial n$ denotes the normal derivative, problem II actually refers only to the given boundary values of φ and $\partial\varphi/\partial n$.

§12. *Supplementary Remarks and Exercises*

1. Variational Problem for a Given Differential Equation. For a given ordinary second order differential equation $y'' = f(x, y, y')$ one can always find a function $F(x, y, y')$ such that the equation $[F]_y = 0$, when solved for y'', is identical with the differential equation.[1]

[1] Cf. O. Bolza, Vorlesungen über Variationsrechnung, Teubner, Leipzig and Berlin, 1909, pp. 37–39.

2. Reciprocity for Isoperimetric Problems. The extremals of the problem

$$J = \int_{x_0}^{x_1} F(x, y, y')\, dx = \text{stationary}$$

under the condition

$$K = \int_{x_0}^{x_1} G(x, y, y')\, dx = \text{constant}$$

are identical with those of the problem $K = \text{stat.}$, $J = \text{const.}$, except for the singular case of §7, 1.

3. Circular Light Rays. The following statement is a consequence of this chapter. Suppose light travels in the x, y-plane in such a way that its speed is proportional to y; then the light rays emitted from any point are circles with their centers on the x-axis.

4. The Problem of Dido. "Dido's problem", to enclose the largest possible area of land within a fence of given length, may be generalized by introducing a weight function $\rho(x, y)$ where $\rho(x, y)$ denotes, say, the fertility of the land. The problem is to find a closed curve of given length such that the integral $\iint_G \rho\, dx\, dy$, taken over the enclosed region, assumes its largest possible value. Formulate the differential equation of the extremals.

5. Examples of Problems in Space. The surface with smallest area which encloses a given volume is the sphere.[1]

Consider the surface of least area bounded by a given curve which, together with a given surface bounded by the same curve, encloses a prescribed volume; then the extremals are the surfaces of constant mean curvature. If the subsidiary condition concerning the volume is omitted, we arrive at the differential equation of *minimal surfaces* (cf. §3, 4), which states that the mean curvature is zero.

6. The Indicatrix and Applications.[2] In finding the extremum of the integral $\int_{t_0}^{t_1} \mathfrak{F}(x, y, \dot{x}, \dot{y})\, dt$ ($\mathfrak{F}$ is assumed positive homogeneous of first order in $\dot{x}$, $\dot{y}$) we consider the curve

$$\mathfrak{F}(x, y, \xi, \eta) = 1$$

in the ξ, η-plane for fixed x, y. This curve is known as the *indicatrix*;

[1] Cf. references to the literature in W. Blaschke, Kreis und Kugel, Leipzig, 1916.

[2] Cf. C. Carathéodory, Über die starken Maxima und Minima bei enfachen Integralen, Math. Ann., Vol. 62, 1906, pp. 449–503.

it enables one to interpret many important relations geometrically. The indicatrix for a three-dimensional problem is the surface with the equation $\mathfrak{F}(x, y, z, \xi, \eta, \zeta) = 1$ in ξ, η, ζ-space.

The direction $(\delta x, \delta y)$ is transverse to $(\dot{x}, \dot{y})$ if $\mathfrak{F}_{\dot{x}}\delta x + \mathfrak{F}_{\dot{y}}\delta y = 0$. But the equation of the tangent to the indicatrix at the point $\dot{x}/\mathfrak{F}, \dot{y}/\mathfrak{F}$ is

$$\left(\xi - \frac{\dot{x}}{\mathfrak{F}}\right)\mathfrak{F}_{\dot{x}} + \left(\eta - \frac{\dot{y}}{\mathfrak{F}}\right)\mathfrak{F}_{\dot{y}} = 0$$

or

$$\xi\mathfrak{F}_{\dot{x}} + \eta\mathfrak{F}_{\dot{y}} = 1.$$

Thus the transverse direction is the direction of the tangent to the indicatrix at the point where the indicatrix intersects the ray joining the origin with the point $(\dot{x}, \dot{y})$. Clearly, transversality is equivalent to orthogonality if and only if the indicatrix intersects all straight lines through the origin at right angles, i.e. if and only if the indicatrix is a circle about the origin. In this case, since $\mathfrak{F}$ is homogeneous, we have $\mathfrak{F}(x, y, \dot{x}, \dot{y}) = \varphi(x, y)\sqrt{\dot{x}^2 + \dot{y}^2}$. If the directions of the extremals and the transverse curves are symmetrically related, a line through the origin O parallel to the tangent to the indicatrix at the point P intersects the indicatrix in a point where the tangent to the indicatrix is parallel to OP.

The indicatrix is particularly useful in the study of *broken extremals*, i.e. extremals which have a corner (discontinuity of slope) at a point x_0, y_0. We wish to know under what circumstances a curve which leads from (x_1, y_1) to (x_0, y_0), arrives there in the direction $(\dot{x}_0^-, \dot{y}_0^-)$, continues in another direction $(\dot{x}_0^+, \dot{y}_0^+)$, and ends at (x_2, y_2) can provide an extremum. In the intervals where the curve has a continuously turning tangent it must satisfy the Euler equations. To investigate the state of affairs at the corner we suppose that the extremal is a member of a family of curves,

$$x(t) + \epsilon\xi(t), \qquad y(t) + \epsilon\eta(t),$$

where $\xi(t)$, $\eta(t)$ are continuously differentiable functions vanishing at the end points. We form the first variation, i.e. we differentiate with respect to ϵ and set $\epsilon = 0$. If we do this for the two segments separately all terms except the boundary terms corresponding to the corner vanish; those at the ends because the end points are kept fixed, and the integrals because of the extremal character of the segments.

Thus we have

$$\xi(t_0)\mathfrak{F}_{\dot{x}}(x_0, y_0, \dot{x}_0^-, \dot{y}_0^-) + \eta(t_0)\mathfrak{F}_{\dot{y}}(x_0, y_0, \dot{x}_0^-, \dot{y}_0^-)$$
$$- \xi(t_0)\mathfrak{F}_{\dot{x}}(x_0, y_0, \dot{x}_0^+, \dot{y}_0^+) - \eta(t_0)\mathfrak{F}_{\dot{y}}(x_0, y_0, \dot{x}_0^+, \dot{y}_0^+) = 0,$$

and, since $\xi(t_0)$ and $\eta(t_0)$ are arbitrary, the "*Weierstrass-Erdmann vertex condition*"

$$\mathfrak{F}_{\dot{x}}(x_0, y_0, \dot{x}_0^-, \dot{y}_0^-) = \mathfrak{F}_{\dot{x}}(x_0, y_0, \dot{x}_0^+, \dot{y}_0^+),$$
$$\mathfrak{F}_{\dot{y}}(x_0, y_0, \dot{x}_0^-, \dot{y}_0^-) = \mathfrak{F}_{\dot{y}}(x_0, y_0, \dot{x}_0^+, \dot{y}_0^+)$$

holds. Therefore, the tangents to the indicatrix at the points at which the indicatrix intersects the vectors $(\dot{x}_0^-, \dot{y}_0^-)$ and $(\dot{x}_0^+, \dot{y}_0^+)$ coincide. The two directions of the curve at the vertex are those of the vectors from the origin to the two tangent points of a double tangent of the indicatrix.

7. Variable Domains. Consider an integral

$$J = \int_{x_0}^{x_1} F(x, u, u')\, dx$$

where not only the function $u(x)$ but also the limits x_0 and x_1 are variable and may depend on the paramter ϵ; then the first variation of the integral contains, in addition to the usual terms, a term due to the variation of the domain. To be specific, the variation is

$$\delta J = \int_{x_0}^{x_1} [F]_u \delta u\, dx + (F_{u'}\delta u + F\delta x)_{x_0}^{x_1} \tag{103}$$

where we have set

$$\delta u = \epsilon\left(\frac{\partial u(x;\epsilon)}{\partial \epsilon}\right)_{\epsilon=0}, \qquad \delta x_1 = \epsilon\left(\frac{\partial x_1(\epsilon)}{\partial \epsilon}\right)_{\epsilon=0}, \qquad \delta x_0 = \epsilon\left(\frac{\partial x_0(\epsilon)}{\partial \epsilon}\right)_{\epsilon=0},$$

and $[F]_u$ is the Euler functional derivative of F.

A similar formula is valid in two (or more) dimensions where the domain of integration may vary with the parameter ϵ. To obtain the variation of the integral

$$J = \iint_G F(x, y, u, u_x, u_y)\, dx\, dy$$

we suppose that the domain G^* (coordinates denoted by x^* and y^*)

which depends on ϵ is mapped onto the original domain G by a transformation

$$\begin{aligned} x^* &= X(x, y; \epsilon) \\ y^* &= Y(x, y; \epsilon). \end{aligned} \tag{104}$$

We assume that this transformation is one-to-one continuously differentiable and reduces to the identity transformation for $\epsilon = 0$. We now assign to the point (x^*, y^*) of G^* a new functional value $u^* = u^*(x^*, y^*; \epsilon)$, which in the old coordinates becomes

$$u^* = u^*\big(X(x, y; \epsilon), Y(x, y; \epsilon)\big) = U(x, y; \epsilon). \tag{104a}$$

Thus our original function $u(x, y)$ represents the surface for $\epsilon = 0$ of the family of surfaces $u^*(x^*, y^*; \epsilon)$—the separate members of which, for fixed ϵ, are given parametrically by the equations (104) and (104a), with x and y as parameters.

We now form the integral

$$J(\epsilon) = \iint_{G(\epsilon)} F[x^*, y^*, u^*(x^*, y^*; \epsilon), u^*_{x^*}(x^*, y^*; \epsilon) u^*_{y^*}(x^*, y^*; \epsilon)]\, dx^*\, dy^*$$

and, introducing x, y as independent variables, transform it into the integral

$$J(\epsilon) = \iint_G F[X, Y, u^*(X, Y; \epsilon), u^*_{x^*}(X, Y; \epsilon), u^*_{y^*}(X, Y; \epsilon)] \frac{\partial(X, Y)}{\partial(x, y)}\, dx\, dy$$

extended over the fixed domain G. The variation is formed by differentiation with respect to ϵ. The following notation is introduced for convenience:

$$\delta x = \epsilon\left(\frac{\partial X}{\partial \epsilon}\right)_{\epsilon=0}, \qquad \delta y = \epsilon\left(\frac{\partial Y}{\partial \epsilon}\right)_{\epsilon=0},$$

$$\delta u = \epsilon\left(\frac{\partial U(x, y; \epsilon)}{\partial \epsilon}\right)_{\epsilon=0} = \epsilon\left(\frac{\partial u^*(X, Y; \epsilon)}{\partial \epsilon}\right)_{\epsilon=0},$$

$$\delta u_x = \epsilon\left(\frac{\partial u^*_{x^*}(x, y; \epsilon)}{\partial \epsilon}\right)_{\epsilon=0}, \qquad \delta u_y = \epsilon\left(\frac{\partial u^*_{y^*}(x, y; \epsilon)}{\partial \epsilon}\right)_{\epsilon=0}.$$

Then

$$\delta J = \iint_G [F_x \delta x + F_y \delta y + F_u \delta u + F_{u_x} \delta u_x + F_{u_y} \delta u_y + F(\delta x)_x + F(\delta y)_y]\, dx\, dy.$$

We can write this integral in a different form if, instead of the variations above, we use the variations at a fixed argument point:

$$\bar{\delta}u = \epsilon\left(\frac{\partial}{\partial\epsilon}\, u^*(x, y; \epsilon)\right)_{\epsilon=0}.$$

Here we have written (x, y) instead of the independent variables (x^*, y^*), and the variation $\bar{\delta}u$ (for which the argument point is fixed) is related to the variation δu (for which the argument point varies with ϵ) by the identity

$$\delta u = \bar{\delta}u + u_x\delta x + u_y\delta y. \tag{105}$$

In the same way we have

$$\delta u_x = (\bar{\delta}u)_x + u_{xx}\delta x + u_{xy}\delta y$$

$$\delta u_y = (\bar{\delta}u)_y + u_{yx}\delta x + u_{yy}\delta y.$$

Introducing these expressions into δJ we then obtain

$$\delta J = \iint_G \{[F]_u\bar{\delta}u + (F_{u_x}\bar{\delta}u)_x + (F_{u_y}\bar{\delta}u)_y + (F\delta x)_x + (F\delta y)_y\}\, dx\, dy$$

or

$$\delta J = \iint_G [F]_u\bar{\delta}u\, dx\, dy + \int_\Gamma \left(F_{u_x}\frac{\partial x}{\partial n} + F_{u_y}\frac{\partial y}{\partial n}\right)\bar{\delta}u\, ds + \int_\Gamma F\left(\delta x\frac{\partial x}{\partial n} + \delta y\frac{\partial y}{\partial n}\right) ds,$$

where n is the outward normal and s the arc length of the boundary Γ of G. Note that in this formula the variation of J is separated, as one might have conjectured, into a part that arises for a fixed domain and a part that can be ascribed to the variation of the domain; the variation of the domain is represented by the normal components of translation vectors (proportional to ϵ) for the boundary points.

8. E. Noether's Theorem on Invariant Variational Problems. Integrals in Particle Mechanics.[1] We consider a family of transformations, depending on a continuous parameter α,

$$\begin{aligned} x^* &= X^*(x, y, u; \alpha) \\ y^* &= Y^*(x, y, u; \alpha) \\ u^* &= U^*(x, y, u; \alpha). \end{aligned} \tag{106}$$

[1] E. Noether, Invariante Variationsprobleme, Nachr. Ges. Göttingen (math.-phys. Kl.), 1918, pp. 235–257.

We suppose that the transformation that corresponds to the value $\alpha = 0$ is the identity. To any surface $u = u(x, y)$ this family of transformations assigns a family of surfaces $u^* = u^*(x^*, y^*; \alpha)$, depending on α, with the parametric representation

$$x^* = X^*(x, y, u(x, y); \alpha) = X(x, y; \alpha),$$
$$y^* = Y^*(x, y, u(x, y); \alpha) = Y(x, y; \alpha),$$
$$u^* = U^*(x, y, u(x, y); \alpha) = U(x, y; \alpha)$$

(the parameters being x, y).

We now suppose that the value of the integral

$$J = \iint_G F(x, y, u, u_x, u_y)\, dx\, dy$$

remains unchanged under the transformations (106), i.e. that for every region G

$$J^* = \iint_{G^*} F(x^*, y^*, u^*, u^*_{x^*}, u^*_{y^*})\, dx^*\, dy^* = \iint_G F\, dx\, dy,$$

where G^* is the domain over which the point (x^*, y^*) varies as (x, y) varies over the domain G. Then we evidently have

$$\delta J = \alpha\left(\frac{\partial J^*}{\partial \alpha}\right)_{\alpha=0} = 0,$$

and from the results of the previous subsection we obtain

$$0 = \iint_G \left\{ [F]_u\, \bar{\delta} u + \frac{\partial}{\partial x}(F_{u_x}\, \bar{\delta} u) + \frac{\partial}{\partial y}(F_{u_y}\, \bar{\delta} u) + \frac{\partial}{\partial x}(F \delta x) + \frac{\partial}{\partial y}(F \delta y) \right\} dx\, dy, \tag{107}$$

Here, as before, we have put

$$\delta x = \alpha\left(\frac{\partial X^*}{\partial \alpha}\right)_{\alpha=0} = \alpha\left(\frac{\partial X}{\partial \alpha}\right)_{\alpha=0},$$
$$\delta y = \alpha\left(\frac{\partial Y^*}{\partial \alpha}\right)_{\alpha=0} = \alpha\left(\frac{\partial Y}{\partial \alpha}\right)_{\alpha=0}, \tag{108}$$

and we also have, in virtue of (105),

$$\bar{\delta} u = \alpha\left(\frac{\partial U}{\partial \alpha}\right)_{\alpha=0} - (U_x + U_u u_x)_{\alpha=0}\, \delta x - (U_y + U_u u_y)_{\alpha=0}\, \delta y. \tag{109}$$

Since equation (107), by hypothesis, is valid for every domain G, the integrand on the right side must vanish, i.e. we must have

$$[F]_u\,\bar{\delta}u + \frac{\partial}{\partial x}(F_{u_x}\,\bar{\delta}u + F\delta x) + \frac{\partial}{\partial y}(F_{u_y}\,\bar{\delta}u + F\delta y) = 0.$$

Analogous formulas are obtained in the case of several dependent variables. For example, if the integral

$$J = \iint_{\mathfrak{G}} F(x,\ y,\ u,\ v,\ u_x,\ v_x,\ u_y,\ v_y)\ dx\ dy$$

is invariant under the continuous transformations

$$x^* = X(x, y, u, v; \alpha), \qquad y^* = Y(x, y, u, v; \alpha),$$
$$u^* = U(x, y, u, v; \alpha), \qquad v^* = V(x, y, u, v; \alpha),$$

one obtains

$$[F]_u\,\bar{\delta}u + [F]_v\,\bar{\delta}v + \frac{\partial}{\partial x}(F_{u_x}\,\bar{\delta}u + F_{v_x}\,\bar{\delta}v + F\,\delta x)$$
$$+ \frac{\partial}{\partial y}(F_{u_y}\,\bar{\delta}u + F_{v_y}\,\bar{\delta}v + F\,\delta y) = 0,$$

where

$$\begin{aligned}\bar{\delta}u &= \alpha\left(\frac{\partial U}{\partial \alpha}\right)_{\alpha=0} - (U_x + U_u u_x + U_v v_x)_{\alpha=0}\,\delta x \\ &\quad - (U_y + U_u u_y + U_v v_y)_{\alpha=0}\,\delta y, \\ \bar{\delta}v &= \alpha\left(\frac{\partial V}{\partial \alpha}\right)_{\alpha=0} - (V_x + V_u u_x + V_v v_x)_{\alpha=0}\,\delta x \\ &\quad - (V_y + V_u u_y + V_v v_y)_{\alpha=0}\,\delta y.\end{aligned} \tag{109a}$$

These results are easily extended to the case of one, or of more than two, independent variables. For one independent variable, a first integral for the extremals

$$F_{u'}\delta u + F_{v'}\delta v + F\delta x = \alpha \cdot \text{const.}$$

can be obtained by integration. Here the expressions δu, δv, δx are functions of x given by (108) and (109a).

For example, we may verify these results for the problem $\int_{x_0}^{x_1} F(u, u')\,dx = \min$. Since the integrand does not contain x ex-

plicitly, it remains invariant under the continuous transformation

$$x^* = x + \alpha, \qquad u^* = u.$$

Thus we obtain for the extremals the integral $F - u'F_{u'} = \text{const.}$—a result which we already know from §4.

If the integral J is invariant under a family of transformations containing n parameters, then we obtain n linearly independent combinations of Euler expressions in the form of divergences, and hence n linearly independent first integrals.

These facts may be illustrated by the *integrals of particle mechanics*. The orbits of the particles of a system of mass points are given by the extremals of the problem

$$\delta J = \delta \int_{t_0}^{t_1} (T - U)\, dt = 0$$

where $T = \frac{1}{2}\sum m(\dot{x}^2 + \dot{y}^2 + \dot{z}^2)$, and where the potential energy depends only on the relative positions of the mass points, i.e. does not change under a translation or rotation of the system as a whole.

This integral is invariant, for example, under the continuous transformations

$$t^* = t, \qquad x^* = x + \alpha, \qquad y^* = y, \qquad z^* = z$$
$$(\text{i.e. } \delta t = \delta y = \delta z = 0, \qquad \delta x = \alpha),$$

or

$$t^* = t, \qquad x^* = x \cos \alpha + y \sin \alpha,$$
$$y^* = -x \sin \alpha + y \cos \alpha, \qquad z^* = z$$
$$(\text{i.e. } \delta t = \delta z = 0, \qquad \delta x = \alpha y, \qquad \delta y = -\alpha x).$$

Hence from the above considerations,

$$T_{\dot{x}} = \sum m\dot{x} = \text{const.}$$
$$yT_{\dot{x}} - xT_{\dot{y}} = \sum m(y\dot{x} - x\dot{y}) = \text{const.}$$

These relations, together with those obtained by permuting x, y, z, express the conservation of *linear* and of *angular momentum*.

If T and U do not contain the time variable t explicitly, the *energy integral* can be obtained in a similar fashion[1] from the fact that the

[1] For more details see E. Bessel-Hagen, Über die Erhaltungssätze der Elektrodynamik, Mathematische Annalen, Volume 84, 1921, pages 258–276.

integral $\int_{t_0}^{t_1} (T - U)\, dt$ is invariant under the transformation $t' = t + \alpha$, $\delta t = \alpha$.

If the integral J remains invariant under transformations which contain an arbitrary function p of the independent variables together with its derivatives up to order k:

$$x^* = X\left(x, y, u, v, p(x, y), \frac{\partial}{\partial x} p(x, y), \cdots, \frac{\partial^k}{\partial y^k} p(x, y)\right),$$

. . . . . ,

then we obtain a linear combination of the Euler expressions and of their total derivatives up to k-th order which vanishes identically; i.e., the Euler equations are not mutually independent.

A simple example is the homogeneous form of the integral

$$J = \int_{t_0}^{t_1} \mathfrak{F}(x, y, \dot{x}, \dot{y})\, dt.$$

The integral remains unchanged when t, $x(t)$, $y(t)$, $\dot{x}(t)$, $\dot{y}(t)$ are replaced by $t(\tau)$, $x(t(\tau))$, $y(t(\tau))$, $dx(t(\tau))/d\tau$, $dy(t(\tau))/d\tau$. Accordingly the Euler expressions $[\mathfrak{F}]_x$, $[\mathfrak{F}]_y$ are connected by the relation

$$\dot{x}[\mathfrak{F}]_x + \dot{y}[\mathfrak{F}]_y = 0.$$

[Cf. formula (31), page 198.][1]

9. Transversality for Multiple Integrals. Suppose the integral

$$\iint_G F(x, y, z, x_u, y_u, z_u, x_v, y_v, z_v)\, du\, dv$$

is to be minimized under the condition that the boundary of the surface $[x(u, v), y(u, v), z(u, v)]$ lie on a given surface $\varphi(x, y, z) = 0$; a formal extension of the procedure applied for curves would lead to the boundary condition

$$\begin{vmatrix} F_{x_u} & F_{x_v} & \varphi_x \\ F_{y_u} & F_{y_v} & \varphi_y \\ F_{z_u} & F_{z_v} & \varphi_z \end{vmatrix} = 0,$$

[1] For a more detailed discussion and for generalizations and applications to mechanics, electrodynamics, and relativity, see the paper by E. Noether referred to above and the references given there.

if we knew that the derivatives concerned existed and were sufficiently regular at the boundary. However, this condition has not been rigorously analyzed or justified.

10. Euler's Differential Expressions on Surfaces. Suppose that the parametric representation of a surface in p, q, r-space is $p = p(\xi, \eta)$, $q = q(\xi, \eta)$, $r = r(\xi, \eta)$, and that the line element on this surface is $ds^2 = e\,d\xi^2 + 2f\,d\xi\,d\eta + g\,d\eta^2$. Then the expression

$$Q[u, u] = \frac{gu_\xi^2 - 2fu_\xi u_\eta + eu_\eta^2}{eg - f^2}$$

is independent of the choice of parameters. The Euler differential expression corresponding to the surface integral

$$\iint_G Q[u, u]\,\sqrt{eg - f^2}\,d\xi\,d\eta$$

is then

$$\Delta u = \frac{\partial}{\partial \xi}\left[\frac{gu_\xi - fu_\eta}{\sqrt{eg - f^2}}\right] + \frac{\partial}{\partial \eta}\left[\frac{-fu_\xi + eu_\eta}{\sqrt{eg - f^2}}\right],$$

and

$$\frac{\Delta u}{\sqrt{eg + f^2}}$$

is independent of the choice of parameters.

11. Thomson's Principle in Electrostatics. Let u, v, w be the components of the electric field intensity in a condenser, i.e. in the region G bounded by two closed surfaces Γ_1, Γ_2 in space. Suppose that the field is free of sources, i.e.

$$u_x + v_y + w_z = 0, \tag{110}$$

and suppose that the charges on the surfaces Γ_1 and Γ_2 are given as $+Q$ and $-Q$, respectively:

$$\begin{aligned} \iint_{\Gamma_1} (ux_n + vy_n + wz_n)\,ds &= Q, \\ \iint_{\Gamma_2} (ux_n + vy_n + wz_n)\,ds &= -Q \end{aligned} \tag{111}$$

(see page 251 for notation).

Then for electrostatic equilibrium the energy of the field, given by

$$\frac{1}{2}\iiint_G (u^2 + v^2 + w^2)\, dx\, dy\, dz$$

(apart from a constant), must be as small as possible. Applying the method of Lagrange multipliers and letting the multipliers of (110) and (111) be $\lambda(x, y, z)$ and μ_1, μ_2, respectively, we obtain the Euler equations

$$u = \lambda_x, \qquad v = \lambda_y, \qquad w = \lambda_z \tag{112}$$

and the natural boundary conditions

$$\begin{aligned} \lambda &= \mu_1 = \text{const.} \quad \text{on } \Gamma_1 \\ \lambda &= \mu_2 = \text{const.} \quad \text{on } \Gamma_2 . \end{aligned} \tag{113}$$

The field vector with components u, v, w is therefore the gradient of a potential λ, which is constant on each surface and satisfies the potential equation $\Delta\lambda = 0$. This result may also be obtained without using the method of multipliers by representing the vector (u, v, w) as the curl of another vector, thus eliminating the subsidiary condition (110).

12. Equilibrium Problems for Elastic Bodies. Castigliano's Principle. A further example for the theory of §11 is the equivalence of Castigliano's principle and the principle of the minimum of potential energy for the equilibrium of an isotropic elastic body. Before formulating the equilibrium conditions for the case of a three-dimensional elastic body, we recall some fundamentals of the mathematical theory of elasticity.

Suppose that at rest the body in question occupies in x, y, z-space a region G with piecewise smooth boundary surfaces Γ. Suppose that some forces deform the body from this rest position into a new equilibrium position, each point (x, y, z) of the body undergoing a small displacement (u, v, w). We define the strains

$$\begin{aligned} &\epsilon_{11} = u_x, && \epsilon_{12} = \tfrac{1}{2}(u_y + v_x), && \epsilon_{13} = \tfrac{1}{2}(u_z + w_x), \\ &\epsilon_{21} = \tfrac{1}{2}(v_x + u_y), && \epsilon_{22} = v_y, && \epsilon_{23} = \tfrac{1}{2}(v_z + w_y), \\ &\epsilon_{31} = \tfrac{1}{2}(w_x + u_z), && \epsilon_{32} = \tfrac{1}{2}(w_y + v_z), && \epsilon_{33} = w_z \end{aligned} \tag{114}$$

and the dilatation

$$\epsilon = \epsilon_{11} + \epsilon_{22} + \epsilon_{33} .$$

The elastic forces produced by the deformation are given by the nine components of the stress tensor

$$\mathfrak{S} = \begin{pmatrix} S_{11} & S_{12} & S_{13} \\ S_{21} & S_{22} & S_{23} \\ S_{31} & S_{32} & S_{33} \end{pmatrix}$$

which are likewise functions of position satisfying the symmetry relations $S_{12} = S_{21}$, $S_{23} = S_{32}$, $S_{31} = S_{13}$.

The stresses and strains are linearly related by Hooke's law

$$\begin{array}{lll} S_{11} = a\epsilon_{11} + b\epsilon, & S_{12} = a\epsilon_{12}, & S_{13} = a\epsilon_{13}, \\ S_{21} = a\epsilon_{21}, & S_{22} = a\epsilon_{22} + b\epsilon, & S_{23} = a\epsilon_{23}, \\ S_{31} = a\epsilon_{31}, & S_{32} = a\epsilon_{32}, & S_{33} = a\epsilon_{33} + b\epsilon, \end{array} \tag{115}$$

where a and b are positive constants of the material.

Suppose a force $\mathbf{P}$ whose components have the densities P_1, P_2, P_3 acts at each point (x, y, z) of the elastic body; suppose, moreover, a surface force $\mathbf{p}$ whose components have the surface densities p_1, p_2, p_3 acts at each point of the surface Γ of the body. Then the equilibrium conditions are

$$\begin{aligned} \frac{\partial S_{11}}{\partial x} + \frac{\partial S_{21}}{\partial y} + \frac{\partial S_{31}}{\partial z} + P_1 &= 0, \\ \frac{\partial S_{12}}{\partial x} + \frac{\partial S_{22}}{\partial y} + \frac{\partial S_{32}}{\partial z} + P_2 &= 0, \\ \frac{\partial S_{13}}{\partial x} + \frac{\partial S_{23}}{\partial y} + \frac{\partial S_{33}}{\partial z} + P_3 &= 0 \end{aligned} \tag{116}$$

or, in vector notation,

$$\operatorname{div} \mathfrak{S} = -\mathbf{P}$$

in the interior; on the boundary they are

$$\begin{aligned} S_{11}x_n + S_{21}y_n + S_{31}z_n - p_1 &= 0, \\ S_{12}x_n + S_{22}y_n + S_{32}z_n - p_2 &= 0, \\ S_{13}x_n + S_{23}y_n + S_{33}z_n - p_3 &= 0 \end{aligned} \tag{117}$$

or, in vector notation,

$$S_n = \mathbf{p}$$

(the subscript n denotes the outer normal on Γ).

The problem now is to determine the stresses and displacements at each point if the following quantities are given: the forces P_1, P_2, P_3 in G, the surface forces $p_1 = \bar{p}_1$, $p_2 = \bar{p}_2$, $p_3 = \bar{p}_3$ on a portion Γ_1 of the boundary, and the displacements

$$u = \bar{u}, \qquad v = \bar{v}, \qquad w = \bar{w} \tag{118}$$

on another part Γ_2 of the boundary.[1]

The equilibrium state is again characterized by the *principle of the minimum of the potential energy*

I: $U[u, v, w] \equiv$

$$\tfrac{1}{2} \iiint_G [\epsilon_{11}S_{11} + 2\epsilon_{12}S_{12} + \epsilon_{22}S_{22} + 2\epsilon_{23}S_{23} + \epsilon_{33}S_{33} + 2\epsilon_{31}S_{31}]\, dx\, dy\, dz$$

$$- \iiint_G (P_1u + P_2v + P_3w)\, dx\, dy\, dz$$

$$- \iint_{\Gamma_1} (p_1u + p_2v + p_3w)\, ds.$$

For the variation of this integral the argument functions are the displacements u, v, w, which attain the prescribed values on Γ_2. The stresses are to be expressed in terms of the strains ϵ in accordance with (115); the strains, in turn, are given in terms of the displacements by relations (114).

The equilibrium conditions (116) in G and (117) on Γ_1 are obtained without difficulty as the Euler equations and natural boundary conditions for U.

If we apply the theory of reciprocal variational problems developed previously to the minimum potential energy principle, we are led to another variational principle, known as "*Castigliano's principle.*" It requires that the work of deformation be a minimum.

To formulate these two reciprocal variational problems in accordance with the theory of §11, we shall employ vector notation and introduce the following definitions: First, we consider two arbitrary symmetric tensors $\mathfrak{S}$ and ϵ connected as above by Hooke's equations (115); if in addition the tensor ϵ is derived from a vector field $\mathbf{s}$ with components u, v, w by the relations (114), i.e. if ϵ is a strain tensor

[1] The normal displacement and tangential force, or tangential displacement and normal force, could also be given on the boundary. It goes without saying that either Γ_1 or Γ_2 may constitute the entire boundary surface.

of a deformation $\mathbf{s}$, we say that $\mathfrak{S}$ satisfies the *compatibility conditions*, or $\mathfrak{S} = \{\mathbf{s}\}$. These conditions, which need not be written out explicitly, characterize $\mathfrak{S}$ as the stress tensor of a deformation $\mathbf{s}$. By using (115) the first integral in our expression for U can be written as

$$\iiint_G \left[a \sum_{i,k} \epsilon_{ik}^2 + b(\epsilon_{11} + \epsilon_{22} + \epsilon_{33})^2 \right] dx\, dy\, dz$$

or, in terms of $\mathfrak{S}$, as $U = Q(\mathfrak{S}, \mathfrak{S})$, where Q is a positive definite quadratic integral, explicitly

$$(119)\quad Q(\mathfrak{S}, \mathfrak{S}) = \iiint_G \frac{1}{a} \left[\sum_{i,k} S_{ik}^2 - \frac{b}{a + 3b} (S_{11} + S_{22} + S_{33})^2 \right] dx\, dy\, dz.$$

For this quadratic form the following Green identity holds: If $\mathfrak{B}$, $\mathfrak{C}$ are two symmetric tensors of which $\mathfrak{C} = \{\mathbf{c}\}$ is a stress tensor for the deformation $\mathbf{c}$, then

$$(120)\quad Q(\mathfrak{B}, \mathfrak{C}) = -\iiint_G \mathbf{c} \operatorname{div} \mathfrak{B}\, dx\, dy\, dz + \iint_\Gamma \mathbf{c}\, \mathfrak{B}_n\, ds.$$

Now we consider two reciprocal linear subspaces Ω and Σ in the space of tensors $\mathfrak{B}$, $\mathfrak{C}$, $\mathfrak{S}$, $\mathfrak{T}$, etc. Ω consists of symmetric tensors $\mathfrak{D}$ for which

$$\operatorname{div} \mathfrak{D} = 0 \quad \text{in } G, \qquad \mathfrak{D}_n = 0 \quad \text{on } \Gamma_1$$

and Σ consists of tensors $\mathfrak{Z}$ for which

$$\mathfrak{Z} = \{\mathbf{s}\} \quad \text{in } G, \qquad \mathbf{s} = 0 \quad \text{on } \Gamma_2.$$

The reciprocity of these two spaces follows from (120).

Now we consider two tensors $\mathfrak{S}^0$, $\mathfrak{T}^0$ subject only to the following conditions:

(a) $$\mathfrak{S}^0 = \{\mathbf{s}^0\} \text{ in } G,$$

where $\mathbf{s}^0$ is an arbitrary strain satisfying $\mathbf{s}^0 = \mathbf{s}$ on Γ_2, and

(b) $$\operatorname{div} \mathfrak{T}^0 = -\mathbf{P} \quad \text{in } G, \qquad \mathfrak{T}_n^0 = \bar{\mathbf{p}} \quad \text{on } \Gamma_1.$$

Furthermore, we call tensors $\mathfrak{S}$ and $\mathfrak{T}$ admissible if $\mathfrak{S} - \mathfrak{S}^0$ is in Σ and $\mathfrak{T} - \mathfrak{T}^0$ is in Ω. Then the problems

I*: $$Q(\mathfrak{S} - \mathfrak{T}^0, \mathfrak{S} - \mathfrak{T}^0) = \min.$$

II*: $$Q(\mathfrak{T} - \mathfrak{S}^0, \mathfrak{T} - \mathfrak{S}^0) = \min.$$

are reciprocal in the sense of §11. By using (120) and by disregarding terms which are a priori known from the data (including $\mathfrak{S}^0$, $\mathfrak{T}^0$), we obtain the following form of the two reciprocal problems:

$$\text{I:} \quad Q(\mathfrak{S}, \mathfrak{S}) - 2 \iiint_G \mathfrak{sP}\, dx\, dy\, dz - 2 \iint_{\Gamma_1} \mathfrak{s}\bar{\mathfrak{p}}\, ds = \min.,$$

which is the original principle of minimum potential energy, and

$$\text{II:} \quad Q(\mathfrak{T}, \mathfrak{T}) - 2 \iint_{\Gamma_2} \bar{\mathfrak{s}}\mathfrak{T}_n\, ds = \min.,$$

which is Castigliano's principle of minimum work of deformation. The normal component $\mathfrak{T}_n$ of the tensor $\mathfrak{T}$ is to be expressed by (117); the variational conditions in G are the compatibility conditions, not written down explicitly.

In these two reciprocal problems one observes again that the variational conditions for one are the admissibility conditions for the other.

It is furthermore remarkable that in the formulations of problems I and II any reference to $\mathfrak{S}^0$ and $\mathfrak{T}^0$ has disappeared as far as the interior of G is concerned. These problems can therefore be formulated without constructing tensors $\mathfrak{S}^0$ and $\mathfrak{T}^0$ (by solving underdetermined partial differential equations).

By Castigliano's principle one can treat certain problems of mechanics more simply than by the principle of minimum potential energy. For example, in the theory of elastic beams the analogue of Castigliano's principle reduces the problem to one for an ordinary minimum. This can be seen in the following way: we may explicitly satisfy the homogeneous differential equation of the beam by functions containing only a finite number of parameters; then the splitting of the function space into orthogonal subspaces leads to spaces Σ which depend only on a finite number of parameters such that problem II reduces to an ordinary minimum problem.

13. The Variational Problem of Buckling. If a rod is compressed by a longitudinal force P acting at both ends, it may be in stable or in unstable equilibrium; i.e., after a slight lateral bending it will either return to the equilibrium position or "buckle," depending on whether the magnitude P is less than or greater than a certain critical value P_0, the "buckling force." In the first case the straight rod is in a position of minimum potential energy with respect to small deformations; in the second case, it is not.

If the rod at equilibrium is of length l and if its lateral displacement is denoted by $u(x)$ $(0 \leq x \leq l)$, then the potential energy is—apart from material constants—

$$U = \int_0^l (u'')^2 \, dx - P \int_0^l (u')^2 \, dx.$$

The first term is the energy of bending, the second term the potential energy of elongation (as in the case of the string).

For sufficiently small values of P the minimum of U with the boundary condition $u(0) = u(l) = 0$ has the value zero.[1] On the other hand, for sufficiently large P, U can be negative; for any admissible function u we need only choose

$$P > \frac{\int_0^l (u'')^2 \, dx}{\int_0^l (u')^2 \, dx}.$$

The buckling force P_0, i.e. the largest value of P for which the minimum of U is equal to zero, may evidently be expressed as the minimum of

$$\frac{\int_0^l (u'')^2 \, dx}{\int_0^l (u')^2 \, dx}$$

with the boundary condition $u(0) = u(l) = 0$. Equivalently, P_0 is the minimum of

$$\int_0^l (u'')^2 \, dx$$

subject to the condition

$$\int_0^l (u')^2 \, dx = 1$$

[1] For example, this is true for $P < 1/l^2$. For, since $\int_0^l u' \, dx = 0$ there is a point x_0 at which $u'(x_0) = 0$. Then

$$u'(x) = \int_{x_0}^x u''(x) \, dx, \quad (u')^2 \leq l \int_0^l (u'')^2 \, dx, \quad \int_0^l (u')^2 \, dx \leq l^2 \int_0^l (u'')^2 \, dx.$$

and the boundary condition $u(0) = u(l) = 0$. This is expressed by saying that $P_0 = \lambda$ is the first eigenvalue of the differential equation

$$u'''' + \lambda u'' = 0$$

with the boundary conditions $u(0) = u(l) = 0$, $u''(0) = u''(l) = 0$. Such eigenvalue problems and their treatment by means of variational methods will be discussed in the next two chapters.

References

For detailed references see the excellent bibliography by Lecat:

Lecat, M. Bibliographie du calcul des variations 1850–1913. Gand, Paris, 1913.

Lecat, M. Bibliographie du calcul des variations depuis les origines jusqu'à 1850 comprenant la liste des travaux, qui ont préparé ce calcul. Gand, Paris, 1916.

Textbooks

Here we mention only the most important textbooks:

Bliss, G. A. Calculus of Variations. Open Court, Chicago, 1924.

Bolza, O. Vorlesungen über Variationsrechnung. B. G. Teubner, Leipzig and Berlin, 1909.

Hadamard, J. Leçons sur le calcul des variations, Vol. I. A. Hermann et fils, Paris, 1910.

Kneser, A. Lehrbuch der Variationsrechnung. F. Vieweg und Sohn, Braunschweig, 1925.

Moigno, M., and Lindelöf, L. L. Calcul des variations. Mallet-Bachelier, Paris, 1861.

Tonelli, L. Fondamenti di calcolo delle variazioni, Vols. I and II. N. Zanichelli, Bologna, 1921 and 1923.

Vivanti, G. Elementi del calcolo delle variazioni. Giuseppe Principato, Messina, 1923.

Monographs and Articles

The following articles from the Enzyklopädie der math. Wiss., 1904–1916:

Hellinger, E. Die allgemeinen Ansätze der Mechanik der Kontinua. Vol. 4D, pp. 601–694.

Kneser, A. Variationsrechnung, Vol. 2A, pp. 571–625.

Zermelo, E., and Hahn, H. Weiterentwicklung der Variationsrechnung in den letzten Jahren, Vol. 2A, pp. 626–641.

French edition of the encyclopedia:

Lecat, M. Calcul des variations, Tome II, Vol. 6, pp. 1–288.

CHAPTER V

Vibration and Eigenvalue Problems

In Chapters V and VI important aspects of the linear differential equations of mathematical physics will be discussed, especially in relation to vibration phenomena. The method of eigenfunctions will occupy a central position throughout.

§1. *Preliminary Remarks about Linear Differential Equations*

1. Principle of Superposition. A linear homogeneous differential expression (or differential operator) for a function u is generally written in the form

$$L[u] = Au + Bu_x + \cdots + Cu_{xx} + \cdots$$

where the coefficients are given functions of the independent variables. $L[u]$ is a linear homogeneous combination of the function u and its derivatives up to a given order, called the *order of the differential expression. The linear differential operator*

$$L = A + B(\partial/\partial x) + \cdots + C(\partial^2/\partial x^2) + \cdots$$

satisfies the fundamental relation

$$L[c_1u_1 + c_2u_2] = c_1L[u_1] + c_2L[u_2] \tag{1}$$

where c_1, c_2 are any constants. The general linear differential equation has the form

$$L[u] = f(x, y, \cdots),$$

where f is a given function of the independent variables. If $f \equiv 0$ the differential equation is called homogeneous, and otherwise nonhomogeneous.

In this chapter we shall be almost exclusively concerned with linear homogeneous differential operators; these operators are a special

instance of linear homogeneous functional operators. Other examples are given by the integral expression

$$\iint_G K(x, y; \xi, \eta)u(\xi, \eta)\, d\xi\, d\eta,$$

by the operator

$$\Theta[u] = \frac{2}{h^2\pi}\int_0^{2\pi} \{u(x + h\cos\theta, y + h\sin\theta) - u(x, y)\}\, d\theta,$$

or by the difference operator

$$\frac{1}{h^2}\{u(x+h,y) + u(x-h,y) + u(x,y+h) + u(x,y-h) - 4u(x,y)\}.$$

If we suppose that u has continuous derivatives up to the second order, the last two expressions tend to the differential operator Δu for $h \to 0$; this can be easily verified. The condition that a linear combination of such linear operators should equal a known function is expressed by a linear functional equation; integral, difference, and differential equations are of this type. Equation (1) expresses the linear homogeneous character of an operator $L[u]$.

The solutions of a linear homogeneous differential equation, and in general of any linear homogeneous functional equation, have the following *superposition property*: If u_1, u_2 are two solutions of the equation and c_1, c_2 are arbitrary constants, then $c_1u_1 + c_2u_2$ is also a solution. More generally, we can combine any number of known particular solutions u_1, u_2, $\cdots$ with constants c_1, c_2, $\cdots$ and obtain a new solution $c_1u_1 + c_2u_2 + \cdots$. A convergent series $\sum_{n=1}^{\infty} c_nu_n$ formed from an infinite sequence of solutions u_1, u_2, $\cdots$ represents a solution if the differential operator $L[u]$ may be applied to the series term by term.

If a solution $u(x, y, \cdots; \alpha)$ of the functional equation $L[u] = 0$, depending on a parameter α, is known, new solutions of the form

$$v = \int w(\alpha)u(x, y, \cdots; \alpha)\, d\alpha$$

can be constructed; here $w(\alpha)$ is an arbitrary function and the domain of integration can be chosen at will. The only restrictions needed are: the integral must exist and the operation L must be applicable under the integral sign. These conditions are, in particular, satisfied for

differential operators if $w(\alpha)$ is piecewise continuous and the domain of integration is finite.

If the homogeneous equation is completely solved one solution of the inhomogeneous equation yields all its solutions; for, all solutions of the inhomogeneous equation are obtained by adding one such solution to each solution of the homogeneous equation.

2. Homogeneous and Nonhomogeneous Problems. Boundary Conditions. We consider problems in which both a linear differential equation and boundary or initial conditions must be satisfied (cf. Ch. IV, §10). If both the boundary conditions and the differential equation are homogeneous, the problem is called homogeneous. In this case, cu is a solution if u is a solution and c is constant. Homogeneous boundary conditions usually consist of relations between the values assumed by the desired function u and its derivatives $u_x, \cdots$ on the boundary Γ of the domain G in question. A simple condition of this kind is $u = 0$ or $\partial u/\partial n = 0$, where $\partial/\partial n$ denotes differentiation in the direction of the outer normal.

Given linear nonhomogeneous boundary conditions –for instance, boundary values $u = f$ (not vanishing everywhere)—we can obtain an equivalent problem with homogeneous boundary conditions: Assume that $L[u] = 0$ is a linear homogeneous equation, and the boundary values f can be extended continuously into the interior of G in such a way that $L[f] = g$ is a continuous function in G. Then, for the new unknown function $v = f - u$, we immediately have the differential equation $L[v] = g$ with the homogeneous boundary condition $v = 0$. Conversely, given a nonhomogeneous linear equation with homogeneous boundary conditions, a special solution of which is known, then a problem with a homogeneous equation and nonhomogeneous boundary conditions can be obtained by subtraction. In general: *Homogeneous differential equations with nonhomogeneous boundary conditions are essentially equivalent to nonhomogeneous differential equations with homogeneous boundary conditions.*[1]

3. Formal Relations. Adjoint Differential Expressions. Green's Formulas. We shall briefly discuss certain formal relations. In particular we shall consider differential expressions which arise from a

[1] However, one should note that the first step, transforming a problem with nonhomogeneous boundary conditions into one with homogeneous conditions, involves assumptions on continuity and differentiability not necessarily made for the original problem.

variational problem with a homogeneous quadratic integrand; such expressions are called *self-adjoint differential expressions.*

(a) *One Independent Variable.* We consider the quadratic expression

$$Q[u, u] = au'^2 + 2bu'u + du^2,$$

where a, b, d are given functions of x, and $u(x)$ is the argument function. We integrate the symmetric bilinear expression

$$Q[u, v] = au'v' + b(u'v + v'u) + duv$$

over an interval (x_0, x_1). Integrating by parts and eliminating the derivatives of v, we obtain *"Green's formula"*

$$\int_{x_0}^{x_1} Q[u, v]\, dx = -\int_{x_0}^{x_1} vL[u]\, dx + (au' + bu)v \Big|_{x_0}^{x_1}, \tag{2}$$

where the differential expression

$$L[u] = (au')' + (b' - d)u$$

agrees, except for the factor -2, with the Euler differential expression associated with the integrand $Q[u, u]$. Similarly, since $Q[u, v]$ is symmetric, we have

$$\int_{x_0}^{x_1} Q[u, v]\, dx = -\int_{x_0}^{x_1} uL[v]\, dx + (av' + bv)u \Big|_{x_0}^{x_1}; \tag{2a}$$

from (2) and (2a) we obtain the symmetric Green's formula

$$\int_{x_0}^{x_1} (vL[u] - uL[v])\, dx = a(u'v - v'u) \Big|_{x_0}^{x_1}. \tag{2b}$$

If we start with an arbitrary bilinear expression

$$B[u, v] = au'v' + bu'v + cuv' + duv$$

instead of a symmetric one and integrate by parts, we obtain formulas of the form

$$\begin{aligned} \int_{x_0}^{x_1} B[u, v]\, dx &= -\int_{x_0}^{x_1} vL[u]\, dx + (au' + cu)v \Big|_{x_0}^{x_1} \\ &= -\int_{x_0}^{x_1} uM[v]\, dx + (av' + bv)u \Big|_{x_0}^{x_1}, \end{aligned} \tag{3}$$

$$\int_{x_0}^{x_1} (vL[u] - uM[v])\, dx = [a(u'v - v'u) + (c - b)uv] \Big|_{x_0}^{x_1}. \tag{4}$$

The differential expression

$$M[v] = (av')' + (bv)' - cv' - dv$$

is associated uniquely with the differential expression

$$L[u] = (au')' - bu' + (cu)' - du$$

by the requirement that the integral on the left in (4) can be expressed in terms of the values of the function and its derivatives on the boundary. These two expressions are said to be *adjoint* to each other. If $L[u] = M[u]$ holds identically, the differential expression $L[u]$ is called *self-adjoint* and may be derived, as above, from a quadratic expression $Q[u, u]$.

The adjoint of the differential expression $pu'' + ru' + qu$ is $(pv)'' - (rv)' + qv$. Hence

$$p' = r$$

is a necessary and sufficient condition for a differential expression to be self-adjoint.

By means of the relations $a = p$, $b' - d = q$, a quadratic expression $Q[u, u]$ that corresponds to $(pu')' + qu$ can be constructed in various ways.

An arbitrary linear differential expression $pu'' + ru' + qu$ can be transformed into one which is self-adjoint: we may multiply by a suitable nonvanishing factor

$$\rho(x) = e^{\int [(r-p')/p]\,dx},$$

we may introduce a new independent variable

$$x' = \int e^{-\int [(r-p')/p]\,dx}\,dx\,.$$

in place of x, or we may introduce a new dependent variable

$$v = ue^{\int [(r-p')/p]\,dx}$$

in place of u.

(b) *Several Independent Variables.* Relations analogous to for mula (2) hold for linear second order partial differential equations An important example is given by the quadratic integrand

$$Q[u, u] = p(u_x^2 + u_y^2) + qu^2$$

with the polar form

$$Q[u, v] = p(u_x v_x + u_y v_y) + quv.$$

We integrate $Q[u, v]$ over a domain G with a piecewise smooth boundary Γ. Integration by parts then yields Green's formula

$$\text{(5)} \qquad \iint_G Q[u, v]\, dx\, dy = -\iint_G vL[u]\, dx\, dy + \int_\Gamma pv \frac{\partial u}{\partial n}\, ds$$

where

$$L[u] = (pu_x)_x + (pu_y)_y - qu.$$

Here we assume that the function v is continuous with piecewise continuous first derivatives in the closed domain G and that u is continuous with continuous first and piecewise continuous second derivatives. In equation (5), s denotes arc length and $\partial/\partial n$ differentiation in the direction of the outer normal.

If v satisfies the same conditions as u, we may interchange u and v in the formula and, subtracting the second expression from (5), obtain the symmetric *Green's Formula*

$$\text{(5a)} \qquad \iint_G (vL[u] - uL[v])\, dx\, dy = \int_\Gamma p\left(v \frac{\partial u}{\partial n} - u \frac{\partial v}{\partial n}\right) ds.$$

For $p = 1$, $q = 0$ our—self-adjoint—differential expression $L[u]$ is the potential expression Δu and formulas (5) and (5a) are the familiar Green's formulas of potential theory.

As in the case of ordinary linear differential equations, an adjoint expression $M[v]$ can be associated with any partial differential expression $L[u]$ if we require that $vL[u] - uM[v]$ can be expressed as a divergence.

4. Linear Functional Equations as Limiting Cases and Analogues of Systems of Linear Equations. Any differential equation can be considered as the limiting case of a difference equation: We replace each differential quotient by a corresponding difference quotient for which the increment of the independent variable, the so-called mesh width, is denoted by h; we consider the functional values of u only at the points $x, y, \cdots$ of a lattice whose coordinates are integral multiples of h. Thus the differential equation is replaced by a system of linear equations for the values of the function u at the lattice points. In a similar way integral equations and other functional

equations can be replaced by systems of linear equations. In Volume II we shall develop this idea systematically, in particular for numerical solutions of differential equations. Here, the analogy between differential equations and difference equations is used merely as a heuristic principle. It leads us to expect that problems in linear differential equations are closely analogous to corresponding problems in the linear algebraic equations from which they originate by limiting processes. This conjecture will be confirmed under very general assumptions.

In particular, the following alternatives hold: If a homogeneous problem corresponding to a homogeneous differential expression has the unique solution $u = 0$, then the nonhomogeneous problem always has one and only one solution. If, however, the homogeneous problem has a nontrivial solution, then the nonhomogeneous problem has solutions only under restrictive linear conditions, and these solutions are not unique. As in Chapter I, the case in which a parameter λ appears linearly in the homogeneous differential expression will play a special role. We are interested in just those values of λ, the *eigenvalues* for the problem, for which the homogeneous problem has a nontrivial solution, an *eigenfunction*.

To replace linear differential equations in the physics of continua by difference equations corresponds to replacing the continuous medium by a system with a finite number of degrees of freedom.

§2. *Systems of a Finite Number of Degrees of Freedom*

As in Chapter IV, §10, we consider a system with n degrees of freedom in the generalized coordinates $q_1, q_2, \cdots, q_n$; the kinetic and potential energies of the system are given by the quadratic forms

$$T = \sum_{h,k=1}^{n} a_{hk}\dot{q}_h\dot{q}_k \qquad \text{and} \qquad U = \sum_{h,k=1}^{n} b_{hk}q_hq_k,$$

respectively, with constant coefficients a_{hk}, b_{hk}.

In view of its physical meaning the form T is positive definite. We shall assume that U is also positive definite. We then know that stable equilibrium occurs for $q_1 = q_2 = \cdots = q_n = 0$.

If we assign nonzero values to some of the coordinates q_h, or if we constrain the q_h by other nonhomogeneous conditions, we obtain a new state of equilibrium different from the original rest state $q_h = 0$. In the case of a finite number of degrees of freedom the problem of

finding such positions of equilibrium under constraints does not introduce a specifically new element into the mathematical formulation; however, in the limit for $n \to \infty$, it leads to the typical boundary value problems of partial differential equations.

1. Normal Modes of Vibration. Normal Coordinates. General Theory of Motion. The motion of the system under consideration is governed by the differential equations

$$\sum_{k=1}^{n} (a_{hk}\ddot{q}_k + b_{hk}q_k) = P_h(t) \qquad (h = 1, 2, \cdots, n) \tag{6}$$

$$(a_{hk} = a_{kh}, \quad b_{hk} = b_{kh})$$

where the functions $P_h(t)$ represent the components of the given external force acting on the system. We seek a solution $q_h(t)$ of this system of differential equations for which the initial position $q_h(0)$ $(h = 1, 2, \cdots, n)$ and the initial velocity $\dot{q}_h(0)$ $(h = 1, 2, \cdots, n)$ are prescribed. If the external forces $P_h(t)$ all vanish, we speak of *free motion* or *free vibration of the system.*

The problem of motion is easily solved by the theory of quadratic forms of Chapter I. The positive definite quadratic forms

$$G = \sum_{h,k=1}^{n} a_{hk}x_hx_k, \qquad F = \sum_{h,k=1}^{n} b_{hk}x_hx_k$$

may be brought into the form

$$G = \sum_{h=1}^{n} \xi_h^2, \qquad F = \sum_{h=1}^{n} \lambda_h\xi_h^2$$

by means of a suitable linear transformation

$$x_h = \sum_{k=1}^{n} \tau_{hk}\xi_k, \qquad \xi_h = \sum_{k=1}^{n} \bar{\tau}_{hk}x_k. \tag{7}$$

Since U and T are definite, the values of $\lambda_1, \lambda_2, \cdots, \lambda_n$ are positive. It follows that in terms of new so-called *normal coordinates* defined by

$$q_h = \sum_{k=1}^{n} \tau_{hk}\eta_k, \qquad \eta_h = \sum_{k=1}^{n} \bar{\tau}_{hk}q_k, \tag{7a}$$

the kinetic and potential energies take the form

$$T = \sum_{h=1}^{n} \dot{\eta}_h^2, \qquad U = \sum_{h=1}^{n} \lambda_h\eta_h^2;$$

the equations of motion take the form

$$\ddot{\eta}_h + \lambda_h \eta_h = N_h(t)$$

where

$$N_h(t) = \sum_{l=1}^{n} P_l(t)\tau_{lh}$$

are *normal coordinates* for the external force. In these differential equations the unknown functions η_h are separated.

It is often convenient to generalize the name "normal coordinate" to include coordinates for which the energies have the form

$$T = c\sum_{h=1}^{n} \dot{\eta}_h^2, \qquad U = \sum_{h=1}^{n} \lambda_h^* \eta_h^2,$$

where $\lambda_h = \lambda_h^*/c = \nu_h^2$.

For free vibrations we have $N_h(t) = 0$, and we immediately obtain the solution in the form

$$\begin{aligned} \eta_h &= y_h \cos \nu_h(t - \varphi_h) \qquad (\nu_h = \sqrt{\lambda_h}) \\ &= a_h \cos \nu_h t + b_h \sin \nu_h t \quad (h = 1, 2, \cdots, n). \end{aligned} \tag{8}$$

Here the a_h, b_h or the y_h, φ_h are arbitrary constants of integration. There is a free motion for which all normal coordinates except the h-th vanish, while the motion of the h-th normal coordinate is given by the equation $\eta_h = y_h \cos \nu_h(t - \varphi_h)$. This motion is called the h-th *normal mode of vibration* or *eigenvibration* of the system with *amplitude* y_h and *phase* φ_h. Whenever we speak of the h-th normal mode of vibration we mean the function $\eta_h = \cos \nu_h t$, i.e. the h-th normal mode with amplitude 1 and phase 0. The numbers ν_i are called the *eigenfrequencies* of the system, or the *pitches*, a term borrowed from acoustics. We obtain the h-th normal mode in terms of the original coordinates q_k if in the transformation formulas (7a) we insert $\cos \nu_h t$ for η_h and 0 for all the other η_i.

Every free motion of the system can be considered as a superposition of eigenvibrations with different phases and amplitudes. The $2n$ constants of integration $a_1, a_2, \cdots, a_n, b_1, b_2, \cdots, b_n$ provide us with the number of arbitrary parameters required to adjust the solution to the prescribed initial state, i.e. to obtain a solution for which the coordinates assume the prescribed initial values and initial velocities.

In order to represent the solution of this initial value problem formally, let us consider the quantities $q_1, q_2, \cdots, q_n$ as components of an n-dimensional vector $\mathfrak{q}$. If we denote the vector whose components are $\tau_{1i}, \tau_{2i}, \cdots, \tau_{ni}$ $(i = 1, 2, \cdots, n)$ by $\mathfrak{e}_i$, we have, by (7a) and (8),

$$\mathfrak{q}(t) = \sum_{i=1}^{n} \mathfrak{e}_i y_i \cos \nu_i(t - \varphi_i).$$

This representation for the general free motion leads immediately to the relations

$$\begin{aligned} \mathfrak{q}(0) &= \sum_{i=1}^{n} \mathfrak{e}_i y_i \cos (\nu_i \varphi_i), \\ \dot{\mathfrak{q}}(0) &= \sum_{i=1}^{n} \mathfrak{e}_i y_i \nu_i \sin (\nu_i \varphi_i), \end{aligned} \tag{9}$$

where the vectors $\mathfrak{q}(0)$, $\dot{\mathfrak{q}}(0)$ describe the given initial state.

If for simplicity we assume that the form T is already the unit form $T = \sum_{i=1}^{n} \dot{q}_i^2$, the "eigenvectors" $\mathfrak{e}_i$ constitute a complete orthogonal system of vectors (cf. Ch. II, §1); multiplying (9) by $\mathfrak{e}_h$ we then obtain the relations

$$\begin{aligned} \mathfrak{e}_h \mathfrak{q}(0) &= y_h \cos (\nu_h \varphi_h), \\ \mathfrak{e}_h \dot{\mathfrak{q}}(0) &= \nu_h y_h \sin (\nu_h \varphi_h), \end{aligned}$$

from which the amplitudes y_h and the phases φ_h can be determined immediately.

We note that the eigenvibrations can be defined as motions of the system for which the ratios of the coordinates q_k are independent of the time, in other words, those for which q_k has the form $q_k = v_k g(t)$, where v_k is independent of t. Substituting these expressions for q_k in equations (6) with $P_i = 0$, we have the equations

$$\frac{\sum_{k=1}^{n} b_{ik} v_k}{\sum_{k=1}^{n} a_{ik} v_k} = -\frac{\ddot{g}(t)}{g(t)}.$$

Since the right member is a constant independent of i and t, say λ, we obtain immediately the eigenvalue problem for G and F formulated in the equations

$$\sum_{k=1}^{n} (b_{ik} - \lambda a_{ik}) v_k = 0 \qquad (i = 1, 2, \cdots, n).$$

The connection with the preceding statements based on transformation (7a) is thus made clear.

We now consider the problem of *forced motion* for which the external forces $P_h(t)$ do not all vanish. To solve this problem it is sufficient to find a single solution of the general differential equation

$$\ddot{\eta}_h + \lambda_h \eta_h = N_h(t).$$

For the initial values $\eta_h(0) = 0$ and $\dot{\eta}_h(0) = 0$ the solution is[1]

$$\eta_h(t) = \frac{1}{\sqrt{\lambda_h}} \int_0^t N_h(\tau) \sin \sqrt{\lambda_h}\,(t - \tau)\, d\tau; \tag{10}$$

and the general forced motion arises from superposition of this special motion on free motion.

Assume that the external force $N_h(t)$ is periodic with frequency ω_h, say of the form $N_h(t) = \alpha_h \cos \omega_h(t - \delta)$. Then if $\omega_h^2 \neq \lambda_h$, formula (10) shows that the motion of the coordinates η_h is the result of superposition of a pure vibration of frequency ω_h and an eigenvibration of frequency $\sqrt{\lambda_h}$. However, if $\omega_h^2 = \lambda_h$, i.e. if "*resonance*" occurs, the forced motion of η_h no longer follows the rhythm of the exciting force $N_h(t)$. Instead, as we see easily from formula (10),

$$\eta_h(t) = \frac{\alpha_h t}{2\omega_h} \sin \omega_h(t - \delta) + \frac{\alpha_h \sin \omega_h \delta}{2\omega_h^2} \sin \omega_h t,$$

and $|\, \eta_h \,|$ no longer remains bounded as t increases.

2. General Properties of Vibrating Systems. We arrange the squares $\lambda_1, \lambda_2, \cdots, \lambda_n$ of the vibration numbers in the order of increasing magnitude: $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$. By Ch. I, §4, we characterize λ_p as the maximum value which can be assumed by the minimum of the quadratic form $F = \sum_{h,k=1}^{n} b_{hk} x_h x_k$ when the variables are subjected first to the condition $G = \sum_{h,k=1}^{n} a_{hk} x_h x_k = 1$, and then to $p - 1$ auxiliary conditions of the form

$$\sum_{h=1}^{n} \alpha_{hj} x_h = 0 \qquad (j = 1, 2, \cdots, p - 1) \tag{11}$$

with arbitrary α_{hj}. From this fact we obtain at once a number of

[1] This solution can be interpreted in the following way: the continuous external force is replaced by discontinuous instantaneous impulses acting at time intervals Δt, and then the limit $\Delta t \to 0$ is taken.

general theorems about the eigenvalues and the corresponding pitches (the proofs have already been given in Ch. I, §4).

THEOREM I: *The p-th overtone of a vibrating system is the highest of the fundamental tones of the systems obtained from the given system by imposing p arbitrarily chosen restrictions of the form* (11).

THEOREM II: *If by imposing r conditions of the form* (11) *on a system S we obtain an "r-fold restricted" system S′, then the frequencies* ν_1', ν_2', $\cdots$, ν_{n-r}' *of the restricted system are neither smaller than the corresponding frequencies* ν_1, ν_2, $\cdots$, ν_{n-r} *nor larger than the frequencies* ν_{r+1}, ν_{r+2}, $\cdots$, ν_n *of the free system; that is,*

$$\lambda_p \leq \lambda_p' \leq \lambda_{p+r}, \qquad \nu_p \leq \nu_p' \leq \nu_{p+r} \quad (p = 1, 2, \cdots, n - r).$$

THEOREM III: *As the inertia increases, the pitch of the fundamental tone and every overtone decreases* (*or remains the same*).

Increase of inertia means change to a system with kinetic energy T' such that $T' - T$ is never negative, while the potential energy remains unchanged.

THEOREM IV: *If the system stiffens, the pitch of the fundamental tone and every overtone increases* (*or remains the same*).

Stiffening of the system means change to a system whose kinetic energy is the same but whose potential energy is greater for the same values of the coordinates.

We need hardly mention that the change in the fundamental tone and overtones is, in each case, opposite to the change indicated in Theorems II to IV if we remove restraining conditions, decrease the masses, or relax the system, i.e. change to a system S' in relation to which S appears stiffened.

§3. *The Vibrating String*

For a system with a finite number of degrees of freedom, the totality of motions is known when the eigenvibrations (synchronous vibrations) are known. This is also true for continuous systems that can vibrate. In such systems we shall consider free vibrations (standing vibrations) for which the deflection u can be expressed as the product of a factor $g(t)$, depending only on the time, and a factor $v(x)$, depending only on the position, *the so-called form factor or vibration form*. An arbitrary vibration phenomenon can be represented by the superposition of eigenvibrations of this kind.

These concepts will be illustrated by a number of examples.

1. Free Motion of the Homogeneous String. The deflection $u(x, t)$ of a homogeneous string with the boundary conditions $u(0, t) = u(\pi, t) = 0$ (cf. Ch. IV, §10) satisfies the differential equation

$$cu_{xx} = \rho u_{tt} \quad \text{or} \quad u_{xx} = \mu^2 u_{tt} \qquad \left(\mu = \sqrt{\frac{\rho}{c}}\right). \tag{12}$$

For simplicity, we choose the unit of time in such a way that $\mu = 1$. We seek functions of the form $u = v(x)g(t)$ which satisfy (12). For these functions the differential equation (12) can be written in the form

$$\frac{v''(x)}{v(x)} = \frac{\ddot{g}(t)}{g(t)}.$$

The left side is independent of t, the right side of x; hence both sides must be equal to one and the same constant, $-\lambda$. From the boundary condition $v(0)g(t) = v(\pi)g(t) = 0$ we have $v(0) = v(\pi) = 0$.

Thus $v(x)$ is to be determined by the differential equation

$$v'' + \lambda v = 0 \tag{13}$$

and the boundary conditions

$$v(0) = v(\pi) = 0. \tag{13a}$$

Not all these requirements can be fulfilled for arbitrary values of the constant λ. In fact, we conclude from the form $c_1 e^{\sqrt{-\lambda}x} + c_2 e^{-\sqrt{-\lambda}x}$ of the general solution of differential equation (13) that the boundary conditions can be fulfilled if and only if $\lambda = n^2$ is the square of an integer n. The corresponding solutions have the form $v_n = \sin nx$. The numbers $1, 2^2, 3^2, \cdots$ and the functions $\sin x, \sin 2x, \cdots$ are called the *eigenvalues* and the *eigenfunctions*, respectively, for the *eigenvalue problem* defined by the differential equation (13) and the boundary conditions (13a).

For $g(t)$ we obtain $g = a \cos nt + b \sin nt$, where a and b are arbitrary constants. Thus for every positive integer n there is a solution of equation (12) of the form $\sin nx(a_n \cos nt + b_n \sin nt)$. The harmonic motions obtained in this way are called the *eigenvibrations* of the string, the corresponding numbers $n = \nu_n$ the associated *eigenfrequencies*. We can form more general solutions by taking sums of the form

$$u = \sum_n \sin nx \, (a_n \cos nt + b_n \sin nt)$$

where the summation extends over either a finite or an infinite number of terms. In the latter case, it is sufficient to assume that the series converges uniformly and that it may be twice differentiated termwise with respect to each of the two variables.

Now we can fit the solution to an arbitrary initial state given by the function $u(x, 0) = \varphi(x)$, $u_t(x, 0) = \psi(x)$ if we choose the coefficients a_n, b_n in a suitable way. For, according to the theory of Fourier series, a_n, b_n can be so determined that

$$\varphi(x) = \sum_{n=1}^{\infty} a_n \sin nx, \qquad \psi(x) = \sum_{n=1}^{\infty} nb_n \sin nx.$$

The series for $u(x, t)$ formed with the coefficients thus determined represents the desired solution.[1]

We obtain quite analogous results if the string is subjected to other boundary conditions. For example, suppose that the initial point is fixed, i.e. $u(0, t) = 0$, and that the end point is connected elastically to its position at rest according to the equation $u_x = -hu$ $(h > 0)$.[2] Then if we set $u(x, t) = v(x)g(t)$ we obtain the following eigenvalue problem for $v(x)$: determine constants $\lambda = \nu^2$ in such a way that the differential equation $v'' + \lambda v = 0$ has a nontrivial solution v satisfying the boundary conditions $v(0) = 0$, $v'(\pi) + hv(\pi) = 0$. The first boundary condition shows that v must have the form $\sin \nu x$. The second boundary condition provides the transcendental equation $h \sin \nu\pi = -\nu \cos \nu\pi$. If $h \neq 0$, we can obtain the roots of this equation graphically by finding in the z, ν-plane the points of intersection of the line $z = -(1/h)\nu$ with the successive branches of the curve $z = \tan \nu\pi$. Thus we again get a sequence of eigenvalues $\lambda_1, \lambda_2, \cdots$ with corresponding eigenfunctions $\sin \nu_1 x$, $\sin \nu_2 x, \cdots$ and eigenvibrations $(a \cos \nu_1 t + b \sin \nu_1 t) \sin \nu_1 x, \cdots$. Moreover, for the n-th eigenfrequency ν_n we obtain immediately the "*asymptotic*" *relation* $\lim_{n\to\infty} (\nu_n/n) = 1$.

For the special case in which the end of the string is "free," that is, in which $h = 0$ and thus $u_x = 0$, we have $\nu_n = n - \frac{1}{2}$, which gives

$$v_n = \sin (n - \tfrac{1}{2})x.$$

[1] These statements are valid if the functions φ, ψ, φ', φ'', ψ' are piecewise smooth. We can relax these assumptions, if we do not require a Fourier expansion of the functions and their derivatives but merely *characterize* them by their Fourier coefficients.

[2] Cf. Ch. IV, §10, 2, where these boundary conditions were derived from the existence of an additional boundary term in the potential energy.

Again we can construct a more general solution of (12) by forming an infinite series

$$u(x, t) = \sum_n \sin \nu_n x(a_n \cos \nu_n t + b_n \sin \nu_n t).$$

By choosing the constants a_n, b_n in a suitable way, we again hope to make this solution satisfy an arbitrarily prescribed initial state. For this purpose, we shall have to examine the *possibility of expanding an arbitrary function* $w(x)$ *in the interval* $0 \leq x \leq \pi$ *in terms of the functions* $\sin \nu_n x$, *the eigenfunctions of the differential equation* (12), under the boundary conditions

$$v(0) = 0, \qquad hv(\pi) = -v'(\pi). \tag{14}$$

This will be done in §14. At this point one should note the *orthogonality property* displayed by the functions $v_n = \sin \nu_n x$, i.e. the property

$$\int_0^\pi v_n v_m \, dx = 0 \quad \text{for} \quad \nu_n \neq \nu_m. \tag{15}$$

This can be verified immediately by multiplying the equation $v_n'' + \nu_n^2 v_n = 0$ by v_m and the equation $v_m'' + \nu_m^2 v_m = 0$ by v_n, subtracting the resulting equations, and integrating. We obtain

$$(\nu_n^2 - \nu_m^2) \int_0^\pi v_n v_m \, dx + \int_0^\pi \frac{d}{dx}(v_n' v_m - v_m' v_n) \, dx = 0,$$

from which the orthogonality property follows by (14).

2. Forced Motion. The motion of a string with fixed end points under the influence of an external force $Q(x, t)$ is characterized by the nonhomogeneous differential equation

$$u_{xx} = u_{tt} - Q(x, t). \tag{16}$$

To find the deflection $u(x, t)$, we expand $Q(x, t)$ at the time t in terms of the eigenfunctions $\sin nx$:

$$Q(x, t) = \sum_{n=1}^{\infty} Q_n(t) \sin nx, \qquad Q_n(t) = \frac{2}{\pi} \int_0^\pi Q(x, t) \sin nx \, dx;$$

similarly, we suppose the desired solution expanded in the form

$$u(x, t) = \sum_{n=1}^{\infty} q_n(t) \sin nx, \qquad q_n(t) = \frac{2}{\pi} \int_0^\pi u(x, t) \sin nx \, dx.$$

We now attempt to satisfy the differential equation (16) by solving the infinite sequence of ordinary differential equations

$$-n^2 q_n(t) = \ddot{q}_n(t) - Q_n(t). \tag{17}$$

Their solutions are given by the functions

$$q_n(t) = \frac{1}{n}\int_0^t Q_n(t') \sin n(t - t')\, dt' + a_n \cos nt + b_n \sin nt \tag{17a}$$

where a_n , b_n are arbitrary constants to be determined from the initial conditions; $\sum_n q_n(t) \sin nx$ is the desired solution of equation (16) if the series converges and is twice termwise differentiable. Another method of handling the nonhomogeneous equation will be developed in §5, 2 and §14, 1.

For forced motion, the result can be obtained without using the expansion theorem. Assume a solution $u(x, t)$ exists; we attempt to determine its Fourier coefficients $q_n(t)$ in terms of the Fourier coefficients $Q_n(t)$ of the function $Q(x, t)$. First we multiply equation (16) by $\sin nx$ and integrate over the fundamental domain. Then, transforming the left side by integration by parts, we obtain (17), and thus we again arrive at formula (17a). The function $u(x, t)$ is characterized uniquely by the expansion coefficients obtained in this way, because the orthogonal system of functions $\sin nx$ is complete.

As in §2, we are particularly interested in the case of harmonic or sinusoidal $Q_n(t)$:

$$Q_n(t) = a \cos \omega t + b \sin \omega t.$$

Here, for $\omega^2 \neq n^2$, $q_n(t)$ can be expressed as a linear combination of a sinusoidal function of frequency ω and one of frequency n, while in the case of resonance, i.e. $\omega^2 = n^2$, $q_n(t)$ becomes unbounded (cf. page 285).

The relations for the homogeneous vibrating string are typical of those for the more general continuous vibrating system with which we shall be concerned in the present chapter. In general, the essential points will be the determination of the normal modes, the question of their completeness, and the validity of the expansion theorem. Here, however, it will not be possible to refer to an existing theory such as the theory of Fourier series. In order not to interrupt the discussion, the completeness proofs for the eigenfunctions will be postponed to §14.

3. The General Nonhomogeneous String and the Sturm-Liouville Eigenvalue Problem. We now consider the general equation of a non-homogeneous string

$$(pu_x)_x = \rho u_{tt}$$

where $p(x)$ denotes the modulus of elasticity multiplied by the cross-sectional area, and $\rho(x)$ the mass per unit length. We seek solutions of this equation which satisfy certain homogeneous boundary conditions. Again we try to find a solution of the form $u = v(x)g(t)$ and obtain

$$(pv')':v\rho = \ddot{g}:g,$$

which can be satisfied only if each side is equal to one and the same constant, say $-\lambda$. For $v(x)$ we then have the ordinary differential equation

$$(pv')' + \lambda\rho v = 0, \tag{18}$$

and g satisfies the differential equation $\ddot{g} + \lambda g = 0$. If we set $\lambda = \nu^2$—it will soon be evident that negative values of λ do not occur —u takes the form

$$u = v(x)(a \cos \nu t + b \sin \nu t);$$

the function $v(x)$ must be determined from the differential equation (18) and from the boundary conditions. As in the special case of the homogeneous string, we are faced with the *eigenvalue problem of determining the "eigenvalues" λ of equation* (18) *for which a nontrivial solution satisfying the boundary conditions exists.* This solution is called the *eigenfunction* belonging to the eigenvalue λ; it is determined except for an arbitrary constant factor. The following types[1] of boundary conditions for the initial and end points are often imposed:

1. $v(0) = 0$ and $v(\pi) = 0$ (string with fixed ends)
2. $h_0v(0) = v'(0)$ and $-h_1v(\pi) = v'(\pi)$ (elastically attached ends)
3. $v'(0) = 0$ and $v'(\pi) = 0$ (free ends)
4. $v(0) = v(\pi)$ and $p(0)v'(0) = p(\pi)v'(\pi)$;

[1] The calculus of variations shows that these include the main types of natural boundary conditions (cf. Ch. IV and VI).

condition 4 can be interpreted as a condition of periodicity if $p(0) = p(\pi)$.

Note that, in accordance with the physical nature of the problem, the functions p and ρ are positive for $0 \leq x \leq \pi$; this we explicitly assume. Furthermore, h_0 and h_1 must be positive if the position of the string at rest is to be a stable equilibrium.[1]

The problem, as formulated, was first attacked by Sturm and Liouville and is therefore called the *Sturm-Liouville eigenvalue problem*. It may be somewhat generalized by considering instead of (18) the differential equation

$$(pv')' - qv + \lambda\rho v = 0, \tag{19}$$

where q is a given continuous function. The differential equations (18) and (19) can be brought into simple normal form by transforming the independent and dependent variables, respectively. After the transformation $z = v\sqrt{\rho}$, for example, equation (19) becomes

$$\frac{d}{dx}(p^*z') - (q^* - \lambda)z = 0 \tag{20}$$

where

$$p^* = \frac{p}{\rho}, \qquad q^* = -\frac{1}{\sqrt{\rho}}\frac{d}{dx}\left(p\frac{d}{dx}\frac{1}{\sqrt{\rho}}\right) + \frac{q}{\rho}.$$

If $q = 0$, equation (19) can be transformed into the form

$$v'' + \lambda\sigma v = 0, \qquad \sigma = \rho p$$

by introducing in place of x the new variable $\xi = \int dx/p(x)$ and then replacing ξ in turn by x.

Another important transformation of the differential equation (19) is given by

$$u = \sqrt[4]{p\rho}\, v, \qquad t = \int_0^x \sqrt{\frac{\rho}{p}}\, dx, \qquad l = \int_0^\pi \sqrt{\frac{\rho}{p}}\, dx. \tag{20a}$$

Then (19) becomes

$$u'' - ru + \lambda u = 0 \tag{19a}$$

where r denotes a continuous function.[2]

[1] Cf. Ch. IV, §10, 2.

[2] Specifically, $r = (f''/f) + (q/\rho)$ where $f = \sqrt[4]{\rho p}$.

To the eigenfunctions v and the (positive) eigenvalues λ of the differential equation (19) correspond eigenvibrations of the string of frequency $\nu = \sqrt{\lambda}$, represented by the functions

$$v(x)(a_\nu \cos \nu t + b_\nu \sin \nu t).$$

Moreover, the eigenfunctions of the Sturm-Liouville problem furnish systems of orthogonal functions; this general property follows from the differential equation. In fact, if λ_n , λ_m are two different eigenvalues and v_n , v_m the corresponding eigenfunctions, we have, as in subsection 1,

$$(\lambda_n - \lambda_m) \int_0^\pi \rho v_n v_m \, dx + \int_0^\pi \frac{d}{dx} (p[v_n' v_m - v_n v_m']) \, dx = 0.$$

Here the boundary conditions imply the vanishing of the second term on the left, and it follows that the functions $\sqrt{\rho} v_i$ are orthogonal, that is,

$$\int_0^\pi \rho v_n v_m \, dx = 0.$$

We may and shall assume that these functions are normalized. In §14 we shall show: *The eigenvalues* λ *of the differential equation* (19) *for given boundary conditions, ordered with respect to magnitude, form a denumerable sequence* λ_1 , λ_2 , λ_3 , $\cdots$, *and the corresponding system of eigenfunctions is a complete orthogonal system.* Moreover: *Every continuous function* $f(x)$ *which has piecewise continuous first and second derivatives and satisfies the boundary conditions of the eigenvalue problem can be expanded in an absolutely and uniformly convergent series*

$$f = \sum_{n=1}^{\infty} c_n v_n , \qquad c_n = \int_0^\pi \rho f v_n \, dx$$

in terms of the eigenfunctions. This expansion theorem makes it possible to fit the solution

$$u(x, t) = \sum_{n=1}^{\infty} v_n(x)(a_n \cos \nu_n t + b_n \sin \nu_n t)$$

to a prescribed initial state.

All eigenvalues of Sturm-Liouville problems (except those with periodic boundary conditions[1]) are *simple*; i.e., no two linearly in-

[1] Here $\lambda = n^2$ $(n = 1, 2, \cdots)$ is a double eigenvalue of $y'' + \lambda y = 0$, with the two eigenfunctions $\sin nx$ and $\cos nx$.

dependent eigenfunctions v, v^* can correspond to the same eigenvalue λ. Indeed, if there were two such eigenfunctions every solution of (19) irrespective of boundary conditions could be expressed in the form $cv + c^*v^*$; thus, every solution would satisfy the homogeneous boundary conditions prescribed for the eigenfunctions. For boundary conditions 1, 2, and 3, this leads to a contradiction since they contain a homogeneous relation between $v(0)$ and $v'(0)$; on the other hand values for $v(0)$ and $v'(0)$ can be arbitrarily prescribed for a solution of (19).

The eigenvalues λ are all positive for $q \geq 0$, $h_0 \geq 0$, $h_1 \geq 0$. We have, in fact,

$$\lambda = \lambda \int_0^\pi \rho v^2 \, dx = -\int_0^\pi [(pv')'v - qv^2] \, dx$$

$$= \int_0^\pi (pv'^2 + qv^2) \, dx - pv'v \Big|_0^\pi$$

where the right-hand side is positive in the case of boundary conditions 1 to 4. *An eigenvalue must be positive if the corresponding eigenfunction is to represent a periodic vibration. Whenever an eigenvalue is negative an aperiodic motion occurs instead of the corresponding normal mode. We shall see later that this can happen only a finite number of times, even if q is negative.*[1]

The *forced motion of the string* could be analyzed in the same way as the homogeneous string in subsection 2. But if the external force $Q(x, t)$ in the nonhomogeneous equation $(pu_x)_x = \rho u_{tt} - Q(x, t)$ is periodic of the form $Q(x, t) = \varphi(x)e^{i\omega t}$,[2] the following procedure[3] is ordinarily used: We write the solution u in the form $u = v(x)e^{i\omega t}$ and obtain the nonhomogeneous equation for $v(x)$, associated with (18),

$$(pv')' + \lambda \rho v = -\varphi(x) \qquad (\lambda = \omega^2).$$

To determine the expansion coefficients

$$\gamma_n = \int_0^\pi \rho v v_n \, dx$$

[1] Cf. Ch. VI, §2.

[2] As always, the use of complex quantities expresses in a simple way that the real and imaginary parts in equations and solutions should be considered separately.

[3] In this connection cf. the algebraic analogue in Ch. I, §3, 6.

of the solution $v(x)$ we multiply our differential equation by $v_n(x)$, integrate over the fundamental domain, transform the first term using integration by parts, and consider the differential equation for v_n. This implies immediately $\gamma_n(\lambda - \lambda_n) = -c_n$; hence

$$\gamma_n = -\frac{c_n}{\lambda - \lambda_n} \qquad \text{where} \qquad c_n = \int_0^\pi \varphi v_n \, dx.$$

This method becomes meaningless in the case of *resonance*, that is, when the frequency $\sqrt{\lambda} = \omega$ of the external force equals one of the eigenfrequencies $\sqrt{\lambda_n} = \omega_n$ and the corresponding coefficient c_n is different from zero.

The case of a general external force $Q(x, t)$ can be reduced to the special case just considered by spectrally decomposing the force $Q(x, t)$ as a function of t, with the help of a Fourier series or a Fourier integral (cf. Ch. II, §§5 and 6).

§4. *The Vibrating Rod*

In the case of the differential equation

$$\frac{\partial^4 u}{\partial x^4} + \frac{\partial^2 u}{\partial t^2} = 0$$

of the *transverse vibrations of a homogeneous rod*, we again consider the eigenvibrations (we limit our discussion to the homogeneous rod for the sake of brevity, since the nonhomogeneous rod offers no aspects beyond those already treated in §3). As before we write $u = v(x)g(t)$ and obtain

$$-\frac{v''''}{v} = \frac{\ddot{g}}{g} = -\lambda,$$

that is,

$$v'''' - \lambda v = 0, \qquad \ddot{g} + \lambda g = 0, \tag{21}$$

where the constant λ must be determined in such a way that four prescribed homogeneous boundary conditions are satisfied at the ends of the rod. Again we assume that the rest position of the rod is the interval $0 \leq x \leq \pi$. We distinguish various types of boundary conditions (cf. Ch. IV, §10):

1. $v''(x) = v'''(x) = 0$ for $x = 0$ and $x = \pi$ (free ends)
2. $v(x) = v''(x) = 0$ for $x = 0$ and $x = \pi$ (supported ends)

3. $v(x) = v'(x) = 0$ for $x = 0$ and $x = \pi$ (clamped ends)

4. $v'(x) = v'''(x) = 0$ for $x = 0$ and $x = \pi$

5. $\left.\begin{aligned} v(0) &= v(\pi), & v'(0) &= v'(\pi), \\ v''(0) &= v''(\pi), & v'''(0) &= v'''(\pi) \end{aligned}\right\}$ (periodicity).

In all these cases the eigenfunctions and eigenvalues can be given explicitly, since we know the general solution of the first differential equation (21). That is, assuming $\lambda \neq 0$[1] and $\sqrt[4]{\lambda} = \nu$ we have

$$v = c_1 \cos \nu x + c_2 \sin \nu x + c_3 e^{\nu x} + c_4 e^{-\nu x}$$

or

$$v = \xi_1 \cos \nu x + \xi_2 \sin \nu x + \xi_3 \cosh \nu x + \xi_4 \sinh \nu x.$$

For $\lambda = 0$ the general solution degenerates into a polynomial of the third degree $v = \xi_1 + \xi_2 x + \xi_3 x^2 + \xi_4 x^3$.

Any suitable set of four homogeneous boundary conditions to which the rod is subjected yields four homogeneous equations of the form $\sum a_{ik}\xi_k = 0$ $(i = 1, 2, 3, 4)$ in the four quantities ξ_1, ξ_2, ξ_3, ξ_4; to obtain a nontrivial solution, we set the determinant $|\, a_{ik} \,| = 0$, which means that we have a transcendental equation for the eigenvalues λ. Every root of this equation furnishes one or more eigenfunctions, which can be taken as normalized. In particular, for the *rod with both ends free*, the transcendental equation for ν becomes

$$\cosh \nu\pi \cos \nu\pi = 1.$$

With the exception of the function $\xi_1 + \xi_2 x$ belonging to the double eigenvalue $\lambda = 0$, the associated eigenfunctions, not yet normalized, are

$$v = (\sin \nu\pi - \sinh \nu\pi)(\cos \nu x + \cosh \nu x) - (\cos \nu\pi - \cosh \nu\pi)(\sin \nu x + \sinh \nu x).$$

The solutions for the *rod clamped at both ends* are obtained by twice differentiating the above solutions for the free rod, since the resulting functions satisfy both the differential equation and the boundary conditions for the clamped rod. Moreover, every eigenfunction for the clamped rod can be obtained in this way; i.e., each such eigen-

[1] It is easily proved, as in §3, that $\lambda \geq 0$.

function corresponds to an eigenfunction of the free rod, as we may show by integrating twice and choosing appropriate constants of integration. The eigenvalues are the positive solutions of the same transcendental equation; the eigenfunctions are given by the expression

$$v = (\sin \nu\pi - \sinh \nu\pi)(-\cos \nu x + \cosh \nu x) - (\cos \nu\pi - \cosh \nu\pi)(-\sin \nu x + \sinh \nu x).$$

The problem of the rod differs from that of the vibrating string in that multiple eigenvalues can occur. For example, in the problem of a rod free at both ends, the two normalized linearly independent eigenfunctions $v = 1/\sqrt{\pi}$ and $v = x\sqrt{3}/\pi^3$ both correspond to the eigenvalue zero. However, in the case of the clamped rod, we lose these two eigenfunctions and their associated eigenvalue $\lambda = 0$ by differentiating twice.

In all the cases considered, the eigenfunctions of equation (21) form an orthogonal system. Again we let λ_n , λ_m be two distinct eigenvalues and v_n , v_m the corresponding eigenfunctions. Integrating by parts twice we obtain

$$(\lambda_m - \lambda_n) \int_0^\pi v_n v_m \, dx = (v_n v_m''' - v_m v_n''' - v_n' v_m'' + v_m' v_n'') \Big|_0^\pi,$$

where the right-hand side vanishes because of the homogeneous boundary conditions. We shall show later (§14) that *the system of eigenfunctions is complete, and that arbitrary functions possessing continuous first and second and piecewise continuous third and fourth derivatives may be expanded in terms of these eigenfunctions.*

The remainder of the theory of the transverse motions of a rod is quite analogous to the theory of the string, and need not be further developed here.

§5. *The Vibrating Membrane*

1. General Eigenvalue Problem for the Homogeneous Membrane. The differential equation $\Delta u = u_{tt}$ of a vibrating homogenous membrane also leads to an eigenvalue problem, but the eigenvalue problem pertains to a partial differential equation. Suppose that the membrane at rest occupies a domain G in the x, y-plane with the boundary Γ; the remaining assumptions and notations are the same as in Ch. IV, §10, 3. We consider first the simplest boundary condi-

tion, $u = 0$; i.e., we consider a *membrane, stretched in a given frame* If we set $u(x, y, t) = v(x, y)g(t)$ we obtain at once the relation

$$\frac{\Delta v}{v} = \frac{\ddot{g}}{g} = -\lambda$$

between the functions $v(x, y)$ and $g(t)$. From this it follows that λ is a constant, which we set equal to ν^2. The function $v(x, y)$ and the constant λ are found by solving the following *eigenvalue problem*: Determine the parameter λ as an "*eigenvalue*" in such a way that there exists a not identically vanishing continuous function $v(x, y)$ in $G + \Gamma$ which has continuous first and second derivatives in G, satisfies the differential equation

$$\Delta v + \lambda v = 0, \tag{22}$$

and vanishes on the boundary; such a function v may be assumed normalized. The eigenvalues λ must be positive, a condition we have already expressed by writing $\lambda = \nu^2$. This fact follows from the equation

$$\iint_G (v_x^2 + v_y^2)\, dx\, dy = -\iint_G v\Delta v\, dx\, dy = \lambda \iint_G v^2\, dx\, dy,$$

obtained by multiplying equation (22) by v and applying Green's formula (cf. §1). Accordingly, the general solution of the equation $\ddot{g}/g = -\lambda = -\nu^2$ has the form $g = a \cos \nu t + b \sin \nu t$; it is a periodic function of the time. The solution

$$u(x, y, t) = v(x, y)(a \cos \nu t + b \sin \nu t)$$

of the equation of vibration then corresponds to an *eigenvibration of frequency* $\nu = \sqrt{\lambda}$.

The existence of eigenvibrations, or more precisely the existence of a denumerably infinite sequence of eigenvalues $\lambda_1, \lambda_2, \cdots$ and corresponding eigenfunctions $v_1(x, y), v_2(x, y), \cdots$, as well as the associated theorems on *completeness* and on *expansions in series*, will be demonstrated in §14. At this point, we note the following orthogonality property: Any two eigenfunctions v_i, v_k that correspond to two distinct eigenvalues λ_i and λ_k are orthogonal; i.e.,

$$\iint_G v_i v_k\, dx\, dy = 0.$$

The proof, following the previous pattern, is based on the formula

$$(\lambda_i - \lambda_k) \iint_G v_i v_k \, dx \, dy = - \iint_G (v_k \Delta v_i - v_i \Delta v_k) \, dx \, dy = 0,$$

derived from (22) with the aid of Green's formula and of the boundary condition $u = 0$.

The motion of a freely vibrating stretched membrane with arbitrarily prescribed initial conditions given by $u(x, y, 0) = f(x, y)$, $u_t(x, y, 0) = g(x, y)$ can again be represented in terms of the eigenfunctions as a series expansion

$$(23) \qquad u(x, y, t) = \sum_{n=1}^{\infty} v_n(x, y)(a_n \cos \nu_n t + b_n \sin \nu_n t),$$

in which the coefficients a_n , b_n are determined by the initial conditions and are given by

$$a_n = \iint_G f(x, y) v_n(x, y) \, dx \, dy, \qquad b_n = \frac{1}{\nu_n} \iint_G g(x, y) v_n(x, y) \, dx \, dy.$$

To derive this result we assume that the series (23) converges and may be differentiated termwise a sufficient number of times.

The situation is similar if the prescribed boundary condition is of the form $\partial u/\partial n = -\sigma u$, which describes an elastic attachment of the membrane to its boundary. Here σ denotes a positive quantity, in general a function of position on the boundary. The eigenvalue problem can be formulated precisely as above, and the solution of the initial value problem is obtained in a similar way by means of the expansion theorem. Here, too, the eigenvalues λ are positive numbers. To show this we multiply equation (22) by v and integrate over G. Using Green's formula (§1) and keeping in mind the boundary condition $\sigma v + \partial v/\partial n = 0$ we obtain at once

$$\lambda = \lambda \iint_G v^2 \, dx \, dy = \iint_G (v_x^2 + v_y^2) \, dx \, dy + \int_\Gamma \sigma v^2 \, ds.$$

The numbers $\nu = \sqrt{\lambda}$ are the frequencies of the corresponding eigenvibrations. Again, the eigenfunctions associated with different eigenvalues λ_i , λ_k are orthogonal.

It is interesting to consider the limiting case, $\sigma = 0$, of the "*free membrane*" (which can be physically realized by a suitable mechanism). Although under all other boundary conditions every eigen-

value is positive, in this case the eigenvalue $\lambda = 0$ exists and has the associated eigenfunction $v(x, y) = \text{const}$.

2. Forced Motion. The forced motion of a membrane, described by the differential equation

$$\Delta u = u_{tt} - Q(x, y, t), \tag{24}$$

can be treated by the method of §3, 2. We can represent both the external force $Q(x, y, t)$ and the desired function $u(x, y, t)$ as the series expansions $Q(x, y, t) = \sum_{n=1}^{\infty} q_n(t)v_n(x, y)$ and $u(x, y, t) = \sum_{n=1}^{\infty} u_n(t)v_n(x, y)$ in terms of the eigenfunctions $v_n(x, y)$ of the freely vibrating membrane; we now determine the coefficients $u_n(t)$ from the differential equations

$$\ddot{u}_n + \lambda_n u_n = q_n .$$

Alternatively, assuming a periodic external force, we can expand Q and u into a Fourier series. We solve equation (24) for the case of a simply periodic force of the form $\varphi(x, y)e^{i\omega t}$ and obtain a function $v(x, y)e^{i\omega t}$. It is seen at once that the function $v(x, y)$ satisfies the differential equation

$$\Delta v + \lambda v = \varphi(x, y) \qquad (\lambda = \omega^2), \tag{25}$$

which can be solved by expanding $v(x, y)$ in an eigenfunction series of the form $v = \sum_{n=1}^{\infty} \gamma_n v_n$. The coefficients are determined as on page 294, and are given by

$$\gamma_n = \frac{c_n}{\lambda - \lambda_n} \qquad (\lambda = \omega^2)$$

where $c_n = \iint_G \varphi v_n \, dx \, dy$.

3. Nodal Lines. In the case of a string or a rod, the points at which an eigenfunction v_n vanishes are of particular interest: these points are called the "*nodal points*" of the associated eigenvibration $v_n e^{i\nu_n t}$. In the case of eigenvibrations of a membrane, we consider *nodal lines*, i.e. the curves $v_n(x, y) = 0$. These nodal lines are the curves along which the membrane remains at rest during eigenvibrations. Various examples will be discussed presently (see also Ch. VI, §6).

4. Rectangular Membrane. The eigenvalue problem for a membrane depends on the choice of a domain and leads to many special questions. We shall discuss some of these questions with reference

to a rectangular membrane which covers the domain $G(0 \leq x \leq a, 0 \leq y \leq b)$. The eigenvalues and eigenfunctions which correspond, respectively, to the boundary conditions $u = 0$ and $\partial u/\partial n = 0$ are known explicitly. In the first case the eigenvalues are the numbers $\lambda = \pi^2(n^2/a^2 + m^2/b^2)$ with $n, m = 1, 2, \cdots$; the associated (not normalized) eigenfunctions are the products $\sin(n\pi x/a) \sin(m\pi y/b)$. In the second case the eigenvalues are the numbers $\lambda = \pi^2(n^2/a^2 + m^2/b^2)$ with $n, m = 0, 1, 2, \cdots$ and the corresponding eigenfunctions are $\cos(n\pi x/a) \cos(m\pi y/b)$. Here $\lambda = 0$ is also an eigenvalue, as has been previously emphasized. (Note that the eigenfunctions for the stretched rectangular membrane can be obtained from those for the free membrane by differentiating with respect to x and y.)

Thus we have obtained *all* the eigenfunctions for the problem. This can be seen from the fact that $\sin(n\pi x/a) \sin(m\pi y/b)$ form a complete orthogonal system of functions in G. For, if there were another eigenfunction corresponding to a new eigenvalue, it would be orthogonal to each of the eigenfunctions already found. If, on the other hand, it corresponded to one of the above eigenvalues, say λ_i, we could subtract a suitable linear combination of the eigenfunctions associated with λ_i and obtain an eigenfunction orthogonal to these (and therefore to all) eigenfunctions. Hence it would be orthogonal to the complete system of eigenfunctions given by the product of sines; any new eigenfunction must, therefore, vanish identically.

Theorems such as the expansion theorems can be reduced to the properties of Fourier series in two variables discussed in Chapter II.

The present example shows that in the case of a membrane *multiple eigenvalues* may very well occur. They always occur if the ratio $a:b$ of the edges of our rectangle is rational, since the equation $n^2/a^2 + m^2/b^2 = n'^2/a^2 + m'^2/b^2$ may have the nontrivial integral solutions m, n, m', n'. If, for example, G represents the square $a = b = \pi$, then $m' = n, n' = m$ is a solution of this kind and the eigenfunctions which correspond to the boundary condition $u = 0$ are $\sin mx \sin ny$ and $\sin nx \sin my$. The question of the multiplicity of an eigenvalue in the case of the square domain is thus reduced to the number-theoretic problem: in how many ways can a number ν^2 be represented as the sum $\nu^2 = n^2 + m^2$ of two squares?[1]

[1] In this connection cf. Dirichlet-Dedekind, Vorlesungen über Zahlentheorie, §68, pp. 164–166, 4th ed., F. Vieweg und Sohn, Braunschweig, 1894.

The nodal lines for the eigenfunctions sin nx sin my are simply lines parallel to the coordinate axes. However, in the case of multiple eigenvalues many other nodal lines can occur. The zeros of the function α sin mx sin $ny + \beta$ sin nx sin my for a square are one ex-

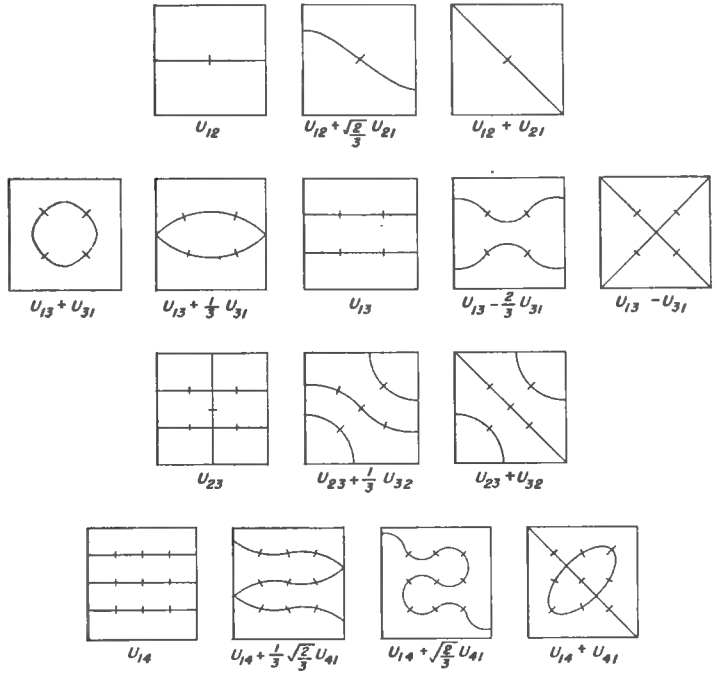


Figure 3. Nodal lines for a square membrane.

ample. The above diagrams[1] illustrate a number of characteristic types. Here u_{mn} denotes the function sin mx sin ny.

5. Circular Membrane. Bessel Functions. The circular membrane can also be treated explicitly; we assume its radius to be 1. The differential equation for its eigenvalue problem, in polar coordinates, takes the form

$$v_{rr} + \frac{1}{r} v_r + \frac{1}{r^2} v_{\theta\theta} + \lambda v = 0, \tag{26}$$

[1] A number are taken from the book by Pockels referred to in the bibliography.

derived in Ch. IV, §8, 2. If, as before, we consider the case of a stretched membrane, we have the boundary condition $v(1, \theta) = 0$. If we attempt to solve equation (26) by setting $v(r, \theta) = f(r)h(\theta)$, we are led at once to the relation

$$\frac{r^2\left(f''(r) + \frac{1}{r}f'(r) + \lambda f(r)\right)}{f(r)} = -\frac{h''(\theta)}{h(\theta)} = \text{const.} = c.$$

Since $v(r, \theta)$ and hence $h(\theta)$ must be periodic functions of θ with period 2π—otherwise v would not be single-valued—it follows that c has a value $c = n^2$, where n is any non-negative integer. We thus obtain

$$h(\theta) = a \cos n\theta + b \sin n\theta$$

and, for $f(r) = y$, the differential equation

$$r^2y'' + ry' + (r^2\lambda - n^2)y = 0 \tag{27}$$

follows. The problem is to find eigenvalues λ for which there exists a solution of this differential equation, continuous for $r = 0$, which also satisfies the boundary condition $f(1) = 0$. If we make the transformation $r\sqrt{\lambda} = \rho$ ($\lambda \neq 0$) or $kr = \rho$, setting $\lambda = k^2$, equation (27) takes the form

$$\frac{d^2y}{d\rho^2} + \frac{1}{\rho}\frac{dy}{d\rho} + \left(1 - \frac{n^2}{\rho^2}\right)y = 0. \tag{28}$$

The solutions of this *Bessel equation*, the so-called *Bessel functions*, play a particularly important role in analysis and mathematical physics and will be the object of more detailed study in Chapter VII. At present, it should be pointed out that by writing y as a power series $y(\rho) = \sum_{m=0}^{\infty} a_m\rho^m$ we obtain for (28) the solution

$$y(\rho) = J_n(\rho)$$

$$= \frac{\rho^n}{2^n n!}\left\{1 - \frac{\rho^2}{2(2n+2)} + \frac{\rho^4}{2\cdot 4(2n+2)(2n+4)} - \cdots\right\}$$

which is called the *Bessel function of n-th order*. The series converges for every value of ρ, as is shown by simple criteria; that is,

the Bessel functions $J_n(\rho)$ are entire transcendental functions. In particular, for $n = 0$ we have the series expansion

$$J_0(\rho) = 1 - \frac{\rho^2}{2^2} + \frac{\rho^4}{2^2 4^2} - \frac{\rho^6}{2^2 4^2 6^2} + \cdots .$$

We note the relation

$$J_0'(\rho) = -J_1(\rho) \tag{29}$$

which follows immediately from the series expansion.

We can write the solutions of (27) in the form

$$y_n = J_n(kr) \qquad (k^2 = \lambda), \tag{30}$$

where the constant k is to be determined by the boundary condition $y_n(1) = 0$, that is, by the condition $J_n(k) = 0$. Thus the eigenvalues $\lambda = k^2$ of (27) are the squares of the zeros of the Bessel functions. As regards the existence of these zeros, we shall show later that each function J_n has, in fact, infinitely many real zeros, which we shall denote by $k_{n,m}$ $(m = 1, 2, 3, \cdots)$. Using this notation we can write the eigenfunctions in the form

$$J_n(k_{n,m}r)(\alpha \cos n\theta + \beta \sin n\theta).$$

The constants α, β in this expression are still arbitrary, which indicates that, with the exception of the eigenfunctions belonging to $n = 0$, all the eigenvalues are at least double since they have the associated linearly independent eigenfunctions $J_n \cos n\theta$ and $J_n \sin n\theta$. The *nodal curves for these eigenfunctions are circles* $\rho = \text{const.}$ *and radial lines* $\theta = \text{const.}$, respectively. The eigenvibrations are represented by

$$u = J_n(k_{n,m}r)(\alpha \cos n\theta + \beta \sin n\theta)(a \cos k_{n,m}t + b \sin k_{n,m}t).$$

If we assume the more general boundary condition $\partial u/\partial r = -\sigma u$, with constant σ, almost all of the above considerations remain unchanged. However, the boundary condition from which the eigenvalues are determined takes a somewhat different form, namely

$$kJ_n'(k) = -\sigma J_n(k).$$

The functions $J_n(k_{n,m}r)$ are the only eigenfunctions for the membrane. This can be proved by noting that every eigenfunction v is a periodic function of θ, with period 2π, having continuous deriva-

tives up to the second order. Hence v can be expanded in a Fourier series

$$v(r, \theta) = \sum_{n=-\infty}^{\infty} f_n(r)e^{in\theta}.$$

Substituting this series in equation (26), we see at once that each individual term $f_n(r)e^{in\theta}$ satisfies the differential equation.

From the general expansion theorem it follows that a function $w(r, \theta)$ vanishing on the boundary and continuous together with its derivatives up to the second order in the interior of the circle, can be expanded in an absolutely and uniformly convergent series of the form

$$w(r, \theta) = \sum_{n,m=0}^{\infty} a_{nm}J_n(k_{n,m}\, r) \cos n(\theta - \theta_{n,m}).$$

For example, if w does not depend on θ, this implies that an arbitrary function of r, which vanishes for $r = 1$ and has continuous derivatives up to the second order in the interval $0 \leq r \leq 1$, can be expanded in this interval in terms of the Bessel functions $J_0(k_{0,m}r)$.

The *orthogonality relation*

$$\int_0^1 rJ_n(k_{n,i}\, r)J_n(k_{n,j}\, r)\, dr = 0 \qquad (i \neq j)$$

for the Bessel functions, or the functions (30), is obtained from the general orthogonality relation for the eigenfunctions of the equation of a membrane by integrating with respect to θ. This relation may also be derived directly from equation (27) using the now familiar method. Furthermore, it is clear that orthogonality is preserved even under the more general boundary condition $kJ'_n(k) = -\sigma J_n(k)$.

To normalize the functions $J_n(k_{n,m}r)$ we shall use relation

$$2\int_0^1 J_n^2(kr)r\, dr = J'^2_n(k), \tag{31}$$

which is proved as follows: We multiply the differential equation for $J_n(kr) = y$, i.e. the equation

$$(ry')' + \left(rk^2 - \frac{n^2}{r}\right)y = 0,$$

by ry', and integrate from 0 to r. Integrating by parts this leads to the relation

$$2k^2 \int_0^r ry^2\,dr = (ry')^2 + (r^2k^2 - n^2)y^2,$$

from which equation (31) follows for $r = 1$, since $y(1) = J_n(k) = 0$.

Thus the functions

$$\frac{\sqrt{2}}{J_n'(k_{n,m})} J_n(k_{n,m}r)$$

are the *normalized eigenfunctions* for equation (27). For further results in the theory of Bessel functions the reader is referred to Chapter VII and to monographs on the subject.

6. Nonhomogeneous Membrane. The generalized differential equation of the nonhomogeneous membrane,

$$p\Delta u + p_x u_x + p_y u_y - qu = \rho(x, y)u_{tt},$$

in which p and ρ are positive in G, leads to an eigenvalue problem analogous to the general Sturm-Liouville problem of §3. This is the problem of determining the values of λ for which the differential equation

$$L[v] + \lambda\rho v = p\Delta v + p_x v_x + p_y v_y - qv + \lambda\rho v = 0$$

possesses a normalized solution satisfying prescribed homogeneous boundary conditions. With the aid of Green's formula

$$\iint_G (v_2 L[v_1] - v_1 L[v_2])\,dx\,dy = \int_\Gamma p\left(v_2 \frac{\partial v_1}{\partial n} - v_1 \frac{\partial v_2}{\partial n}\right) ds = 0$$

[page 280, formula (5a)], we obtain the orthogonality relation

$$\iint_G \rho v_i v_j\,dx\,dy = 0$$

for the eigenfunctions v_i, v_j corresponding to distinct eigenvalues λ_i, λ_j. We should like to determine the eigenfunctions in such a way that the functions $\sqrt{\rho}\,v_i$ form an orthonormal system, i.e. so that

$$\iint_G \rho v_i^2\,dx\,dy = 1.$$

In §14 the *existence of the eigenvalues and the completeness and expansion theorems* will be treated from a general point of view. These

theorems assert that a function $f(x, y)$ which satisfies the boundary conditions and has continuous derivatives of first and second order can be expanded: $f = \sum_{n=1}^{\infty} c_n v_n(x, y)$, with $c_n = \iint_{\blacksquare} \rho f v_n \, dx \, dy$.

§6. *The Vibrating Plate*

1. General Remarks. For the differential equation

$$\Delta\Delta u + u_{tt} = 0$$

of the homogeneous vibrating plate, we obtain the eigenvalue equation

$$\Delta\Delta v - \lambda v = 0, \tag{32}$$

writing $u = v(x, y)g(t)$ with $g(t) = \alpha e^{\pm i\nu t}$ or $g(t) = a \cos \nu t + b \sin \nu t$, and $\lambda = \nu^2$. As boundary conditions we consider for example

$$u = 0, \quad \frac{\partial u}{\partial n} = 0; \quad \text{i.e.,} \quad v = 0, \quad \frac{\partial v}{\partial n} = 0$$

(*clamped plate*). The *orthogonality of two eigenfunctions corresponding to two distinct eigenvalues* is proved by the same method as before using Green's formula (§1). The only essential difference is that now two homogeneous boundary conditions characterize the eigenvalue problem; this corresponds to the fact that the partial differential equation under consideration is of fourth order.

2. Circular Boundary. The problem for the plate is analytically more difficult than that for the membrane. It is not possible, for example, to treat the case of the rectangular boundary in terms of functions known explicitly. The only boundary which has been explicitly treated is the circle. Introducing polar coordinates r, θ we are once more led to Bessel functions. If we set $\lambda = k^4$, we can write the differential equation in symbolic form:

$$(\Delta\Delta - k^4)v = 0$$

or

$$(\Delta - k^2)(\Delta + k^2)v = 0;$$

the operator Δ is given by

$$\Delta = \frac{\partial^2}{\partial r^2} + \frac{1}{r}\frac{\partial}{\partial r} + \frac{1}{r^2}\frac{\partial^2}{\partial \theta^2}.$$

If we suppose that v is expanded in a Fourier series,

$$v = \sum_{n=-\infty}^{\infty} y_n(r)e^{in\theta},$$

then each term of the series must satisfy the differential equation; in other words, y_n must be a solution of

$$\left(\frac{d^2}{dr^2} + \frac{1}{r}\frac{d}{dr} - \frac{n^2}{r^2} - k^2\right)\left(\frac{d^2}{dr^2} + \frac{1}{r}\frac{d}{dr} - \frac{n^2}{r^2} + k^2\right)y = 0.$$

We can find two linearly independent solutions of this differential equation which are regular for $r = 0$: $J_n(kr)$ and $J_n(ikr)$, where $i = \sqrt{-1}$. Thus the function

$$v(r, \theta) = J_n(kr)(a_1 \cos n\theta + b_1 \sin n\theta) + J_n(ikr)(a_2 \cos n\theta + b_2 \sin n\theta)$$

is a solution of (32). To satisfy the boundary conditions $v(1, \theta) = 0$, $v_r(1, \theta) = 0$, we must have

$$J_n(k)a_1 + J_n(ik)a_2 = 0, \qquad J_n(k)b_1 + J_n(ik)b_2 = 0,$$
$$J'_n(k)a_1 + iJ'_n(ik)a_2 = 0, \qquad J'_n(k)b_1 + iJ'_n(ik)b_2 = 0.$$

Hence, the eigenfrequency k satisfies the transcendental equation

$$\frac{J'_n(k)}{J_n(k)} = \frac{iJ'_n(ik)}{J_n(ik)}$$

in which the imaginary unit i no longer really occurs, as is shown by the series expansions of the Bessel functions. For details the reader is again referred to the literature.

§7. *General Remarks on the Eigenfunction Method*

Let us now examine the essential features of the method illustrated by the foregoing examples.

1. Vibration and Equilibrium Problems. Let G be a domain of the independent space variables $x, \cdots$, i.e. an interval on the x-axis or a domain either in the x, y-plane or in x, y, z-space which has a piecewise smooth boundary Γ. Suppose the state of a continuum filling G is characterized by a function $u(x, \cdots ; t)$ which vanishes identically if the system is in stable equilibrium. Let $L[u]$ be a self-adjoint linear differential expression in the independent variables

$x, \cdots$, defined in G, which arises from the variation of the system's potential energy. Let $\rho(x, \cdots)$ represent the mass density at any point of G and let $Q(x, \cdots ; t)$ represent a given external force. We wish to find a solution of the differential equation

$$L[u] = \rho u_{tt} - Q \tag{33}$$

which satisfies prescribed homogeneous time-independent boundary conditions on the boundary Γ of G, and which corresponds to a prescribed initial state defined by

$$u(x, \cdots ; 0) = \varphi(x, \cdots), \qquad u_t(x, \cdots ; 0) = \psi(x, \cdots).$$

All occurring derivatives are assumed to be continuous in G.

The case of equilibrium corresponds to the special assumption that all functions considered are independent of t (and no initial conditions are given). Instead of a mixed initial-boundary value problem for vibration, we then obtain a boundary value problem for equilibrium.

Among the free motions characterized as the solutions of the homogeneous differential equation

$$L[u] = \rho u_{tt} \tag{33a}$$

which satisfy the prescribed homogeneous boundary conditions, we distinguish the eigenvibrations by requiring *synchronism*: $u = v(x, \cdots)g(t)$. Each such eigenvibration is associated with a constant value λ, an eigenvalue, with the property $\ddot{g} + \lambda g = 0$; hence

$$g(t) = a \cos \sqrt{\lambda}\, t + b \sin \sqrt{\lambda}\, t,$$

and

$$L[v] + \lambda \rho v = 0, \tag{34}$$

where v must satisfy the boundary conditions for u given above. The *eigenvalue problem* is to determine values of the parameter λ (eigenvalues) for which the homogeneous differential equation (34) has nontrivial solutions (eigenfunctions) under the prescribed boundary conditions. The vibration satisfying the original equation (33a) is then represented by

$$u = (a \cos \sqrt{\lambda}\, t + b \sin \sqrt{\lambda}\, t)\, v(x, \cdots).$$

In the case of a finite domain G the following statements are in general true: The eigenvalues λ *form a denumerably infinite sequence*

$\lambda_1, \lambda_2, \cdots$. There exists a *system of associated eigenfunctions* $v_1, v_2, \cdots$ which is *complete* in the sense of Ch. II, §1 and which satisfies the orthogonality relations[1]

$$\int_G \rho v_i v_k \, d\tau = 0 \qquad (i \neq k),$$

$$\int_G \rho v_i^2 \, d\tau = 1.$$

Moreover, the *expansion theorem* holds: Every function w with continuous $L[w]$ which satisfies the prescribed homogeneous boundary conditions may be expanded in an absolutely and uniformly convergent series in the eigenfunctions:

$$w = \sum_{\nu=1}^{\infty} c_\nu v_\nu, \qquad c_\nu = \int_G \rho w v_\nu \, d\tau.$$

On the basis of these properties—which must be proved for each problem (cf. §14)—we obtain an infinite sequence of eigenvibrations $(a_\nu \cos \sqrt{\lambda_\nu}\, t + b_\nu \sin \sqrt{\lambda_\nu}\, t) v_\nu(x, \cdots)$. From these eigenvibrations we obtain the solutions of the initial value problem for (33a) by superposition if we choose the constants a_ν and b_ν suitably:

$$a_\nu = \int_G \rho \varphi v_\nu \, d\tau, \qquad b_\nu = \frac{1}{\sqrt{\lambda_\nu}} \int_G \rho \psi v_\nu \, d\tau.$$

For the nonhomogeneous equation (33) with homogeneous boundary conditions—as we saw in §§1, 2, there is no loss of generality in assuming homogeneous boundary conditions for a nonhomogeneous equation—the solution $u(x, \cdots ; t)$ is found by determining its expansion coefficients $\gamma_\nu(t)$ with respect to the v_ν. To this end we multiply equation (33) by v_ν, integrate over G, and transform the left side by Green's formula (5a), §1, taking the boundary conditions into consideration; using (34) we obtain

$$\ddot{\gamma}_\nu + \lambda_\nu \gamma_\nu = Q_\nu(t),$$

where $Q_\nu(t)$ is the given expansion coefficient of $Q(x, \cdots ; t)\rho^{-1}$ with respect to the v_ν. A special solution of this equation for γ_ν is given by

$$\gamma_\nu = \frac{1}{\sqrt{\lambda_\nu}} \int_0^t Q_\nu(\tau) \sin \sqrt{\lambda_\nu}(t - \tau) \, d\tau.$$

[1] The notation $\int_G f \, d\tau$ means the integral of the function $f(x, \cdots)$ over the domain G.

The function formed using these expansion coefficients is a particular solution of (33), and all other solutions are obtained by adding a solution of (33a). Thus the initial value problem under consideration is reduced to the problem of solving the homogeneous equation (33a).

In terms of eigenfunctions we can also solve the equilibrium problem, i.e. the boundary value problem for the differential equation

$$L[u] = -Q(x, \cdots)$$

with homogeneous boundary conditions. In the same way we obtain the equation $\lambda_\nu \gamma_\nu = Q_\nu$ for the expansion coefficients of the desired solution u with respect to the v_ν, i.e.

$$\gamma_\nu = \frac{1}{\lambda_\nu} \int_G Q v_\nu \, d\tau .$$

Hence, by the expansion theorem, the solution is given by

$$u = \sum_{\nu=1}^{\infty} \frac{v_\nu}{\lambda_\nu} \int_G Q v_\nu \, d\tau .$$

If we could interchange summation and integration in this expression we would arrive at a function

$$K(x, \cdots; \xi, \cdots) = \sum_{\nu=1}^{\infty} \frac{v_\nu(x, \cdots) v_\nu(\xi, \cdots)}{\lambda_\nu}$$

such that the solution of the boundary value problem could be written in the form

$$u(x, \cdots) = \int_G Q(\xi, \cdots) K(x, \cdots; \xi, \cdots) \, d\tau,$$

where the integration is to be carried out with respect to the variables $\xi, \cdots$. This function K, *"Green's function"* of $L[u]$, will be characterized in §14 in a quite different manner and will form the basis for a more detailed investigation reaching beyond the formal structure of the present method.

2. Heat Conduction and Eigenvalue Problems. The theory of heat conduction also leads to eigenvalue problems. If the units of time and length are suitably chosen, the differential equation of heat conduction takes the form

$$L[u] = u_t ,$$

where u denotes the temperature as a function of position (x, y, z) and time t. The radiation of heat from a homogeneous body G with surface Γ into an infinite medium of constant temperature zero is characterized at the surface Γ by a boundary condition of the form

$$\frac{\partial u}{\partial n} + \sigma u = 0$$

where σ is a positive physical constant. This condition states that the rate of change of temperature in the direction of the inner normal is proportional to the jump in temperature from the exterior to the interior of the body. We seek a solution of the equation of heat conduction which satisfies this boundary condition and prescribed initial conditions at the time $t = 0$.

We write u in the form $u = v(x, y, z)g(t)$ and obtain at once the equation

$$\frac{L[v]}{v} = \frac{\dot{g}}{g} = -\lambda.$$

We thus have the following eigenvalue problem for v: $L[v] + \lambda v = 0$ in G and $\partial v/\partial n + \sigma v = 0$ on the surface Γ; for a given eigenvalue λ and its eigenfunction v the corresponding solution of the differential equation has the form

$$u = ave^{-\lambda t}.$$

By the eigenfunction expansion theorem we can again make the solution satisfy a given initial state; then $u(x, y, z; 0)$ equals an arbitrarily prescribed function $\varphi(x, y, z)$, which is continuous in G together with its derivatives of first and second order and satisfies the boundary condition. For, if the normalized eigenfunctions $v_1, v_2, \cdots$ and their associated eigenvalues $\lambda_1, \lambda_2, \cdots$ form a complete system, the desired solution is given by the formula

$$u(x, y, z; t) = \sum_{n=1}^{\infty} c_n v_n(x, y, z) e^{-\lambda_n t}$$

where $c_n = \iiint_G \varphi v_n \, dx \, dy \, dz.$

Incidentally, the positive character of the eigenvalues λ implies that the solution $u(x, y, z; t)$ approaches zero asymptotically as t increases as must be expected from the physical meaning of the problem.

If, in place of the homogeneous equation of heat conduction, we consider the inhomogeneous equation

$$L[u] = u_t - Q(x, y, z)$$

(in which we assume that the given function Q does not depend on the time) and impose the same boundary conditions on u as above, we obtain by our general method a solution $u(x, y, z; t)$ which, as $t \to \infty$, goes over into the solution of the corresponding boundary value problem for the equation

$$L[u] = -Q(x, y, z).$$

§8. *Vibrations of Three-dimensional Continua. Separation of Variables*

In the theory of vibrations of three-dimensional continua, e.g. in acoustics, elasticity, or electrodynamics, homogeneous boundary value problems arise for the equation

$$\Delta u = u_{tt}$$

where Δu is the potential expression in three variables; we are led to eigenvalue problems of the form

$$\Delta u + \lambda u = 0$$

with corresponding homogeneous boundary conditions.

It often happens that the particular shape of the fundamental domain permits a further separation of variables for the solutions of this eigenvalue problem and thus leads to new eigenvalue problems involving fewer independent variables.

An example is given by a cylindrical domain erected over the domain G of the x, y-plane and bounded by the planes $z = 0$ and $z = \pi$. Let us take $u = 0$ as a boundary condition. By means of the substitution $u = f(z)v(x, y)$ this problem can be reduced immediately to the corresponding problem for the plane domain G, and we obtain

$$-\frac{f''}{f} = \frac{\Delta v}{v} + \lambda = k = \text{const.},$$

$$f = \sin \sqrt{k}\, z$$

where $k = 1^2, 2^2, 3^2, \cdots$. The equation for v is $\Delta v + (\lambda - n^2)v = 0$; here the eigenvalues differ from those for the plane domain G only

by the terms $-n^2$, and the eigenfunctions are identical with those for the plane domain G.

From the completeness of the system of eigenfunctions, it follows, as in previous cases, that we thus obtain all the eigenfunctions of the cylinder under the given boundary conditions.

If the cylinder is erected over a rectangular domain, we have a rectangular parallelepiped; e.g., for the cube $0 \leq x, y, z \leq \pi$, we obtain the eigenvalues $l^2 + m^2 + n^2$ ($l, m, n = 1, 2, 3, \cdots$) and the eigenfunctions $\sin lx \sin my \sin nz$.

As a further example, we consider the *vibration equation for the spherical domain* $x^2 + y^2 + z^2 \leq 1$ of radius 1. If we introduce polar coordinates r, θ, φ, the vibration equation takes the form (cf. Ch. IV, §8, 2)

$$\Delta u + \lambda u = \frac{1}{r^2 \sin\theta}\left[\frac{\partial}{\partial r}(r^2 u_r \sin\theta) + \frac{\partial}{\partial \varphi}\left(\frac{u_\varphi}{\sin\theta}\right) + \frac{\partial}{\partial\theta}(u_\theta \sin\theta)\right] + \lambda u = 0.$$

If we look for a solution of the form $u = Y(\theta, \varphi)f(r)$, we obtain

$$\frac{(r^2 f')' + \lambda r^2 f}{f} = -\frac{1}{Y \sin\theta}\left[\frac{\partial}{\partial\varphi}\left(\frac{Y_\varphi}{\sin\theta}\right) + \frac{\partial}{\partial\theta}(Y_\theta \sin\theta)\right] = k,$$

where k is a constant. The value of k must be determined in such a way that the differential equation

$$\Delta^* Y + kY = \frac{1}{\sin\theta}\left[\frac{\partial}{\partial\varphi}\left(\frac{Y_\varphi}{\sin\theta}\right) + \frac{\partial}{\partial\theta}(Y_\theta \sin\theta)\right] + kY = 0$$

has a solution which is continuous on the entire surface of the sphere. Therefore, this solution must be periodic in φ, of period 2π, and regular at $\theta = 0$ and $\theta = \pi$ (i.e., at both these points it must approach a limit independent of φ). We shall see in Ch. VII, §5 that this requirement is satisfied only for the values $k = n(n+1)$ ($n = 0, 1, 2, \cdots$); in this case the solutions are the *spherical harmonics* $Y_n(\theta, \varphi)$ (see also §9). We have for $f(r)$ the equation

$$(r^2 f')' - n(n+1)f + \lambda r^2 f = 0;$$

the solutions of this equation which are regular at $r = 0$ are the functions

$$S_n(\sqrt{\lambda}\, r) = \frac{J_{n+\frac{1}{2}}(\sqrt{\lambda}\, r)}{\sqrt{r}}$$

(cf. §5 and §10). We now have to determine the parameter λ from the boundary condition. For example, if the boundary condition is given by $u = 0$, then λ is determined by the equation $J_{n+\frac{1}{2}}(\sqrt{\lambda}) = 0$; denoting the roots of this equation by $\lambda_{n,1}, \cdots$, we obtain solutions of the form $u = Y_n(\theta, \varphi)S_n(\sqrt{\lambda_{n,k}}r)$ for the boundary value problem. In Ch. VII, §5 we shall prove that these solutions constitute a complete orthogonal system, and hence represent all the eigenfunctions and eigenvalues for our differential equation problem.

§9. *Eigenfunctions and the Boundary Value Problem of Potential Theory*

The boundary value problem of potential theory is the problem of determining a function u which satisfies the differential equation $\Delta u = 0$ in the interior of a domain G, and assumes prescribed values on the boundary. In §§1, 2, it was pointed out that we may instead solve the nonhomogeneous equation $\Delta u = f$ with the associated boundary condition $u = 0$. We could investigate the latter problem by the method of §7, expanding f and u in terms of the eigenfunctions for the equation

$$\Delta v + \lambda v = 0.$$

However, for suitable special domains G we can proceed more simply: by separating the variables one can reduce the number of independent variables in the problem. This procedure will be illustrated by a number of important examples.

1. Circle, Sphere, Spherical Shell. Consider first the case of two independent variables x, y; let G be the circle of radius 1 about the origin. If we transform the expression Δu to polar coordinates r, φ, we have the following boundary value problem: to solve the equation

$$r(ru_r)_r + u_{\varphi\varphi} = 0$$

with preassigned boundary values $u(1, \varphi) = f(\varphi)$; here $f(\varphi)$ is a continuous periodic function of period 2π with a piecewise continuous first derivative. If we attempt to find solutions of the homogeneous equation—for the moment ignoring the boundary condition—which can be written in the form $u = v(r)w(\varphi)$, we are led in the usual manner to an eigenvalue problem

$$w'' + \lambda w = 0;$$

the boundary conditions are the periodicity conditions $w(0) = w(2\pi)$, $w'(0) = w'(2\pi)$. The eigenvalues for this problem are given by $\lambda = n^2$ (integral n), the associated eigenfunctions are given by $w = a_n \cos n\varphi + b_n \sin n\varphi$. For $v(r)$ we obtain the differential equation $r(rv')' - n^2 v = 0$, of which $v = r^n$ and $v = r^{-n}$ are linearly independent solutions. Thus we obtain particular solutions of the original equation, which are regular in the unit circle and have the form

$$(a_n \cos n\varphi + b_n \sin n\varphi)r^n,$$

where the constants a_n and b_n are arbitrary. (These solutions can also be characterized as solutions of the differential equation $\Delta u = 0$ which are integral rational functions of x and y and homogeneous of the n-th degree.)

According to the theory of Fourier series we can obtain the desired solution of the boundary value problem by the process of superposition,

$$u = \sum_{n=0}^{\infty} r^n(a_n \cos n\varphi + b_n \sin n\varphi),$$

with the coefficients a, b taken from the series expansion of the preassigned boundary values (cf. Ch. IV, §2, p. 179).

The situation is quite similar in three dimensions, where we consider the unit sphere $x^2 + y^2 + z^2 \leq 1$ for G; then *Laplace's spherical harmonics* occur instead of the trigonometric functions. In fact, if we transform the differential equation to polar coordinates r, θ, φ (cf. pages 225 and 314) we obtain the equation

$$(r^2 u_r)_r + \frac{1}{\sin^2\theta} u_{\varphi\varphi} + \frac{1}{\sin\theta}(u_\theta \sin\theta)_\theta = 0,$$

which for $u = v(r)Y(\theta, \varphi)$ leads to the equation

$$(r^2 v')' - \lambda v = 0. \tag{35}$$

The general solution of (35) has the form

$$v = c_1 r^{\alpha_1} + c_2 r^{\alpha_2}$$

where c_1 and c_2 are arbitrary constants and α_1, α_2 denote roots of the quadratic equation

$$\alpha(\alpha + 1) = \lambda.$$

For Y we obtain the same eigenvalue problem as in §8, given by the differential equation

$$(36)\quad \Delta^* Y + \lambda Y = \frac{1}{\sin\theta}\left[\frac{1}{\sin\theta}\, Y_{\varphi\varphi} + (Y_\theta \sin\theta)_\theta\right] + \lambda Y = 0$$

where λ must be determined in such a way that the equation has a nontrivial solution with continuous derivatives up to the second order over the entire sphere.

To justify our requirement that $Y(\theta, \varphi)$ be regular even at the poles $\theta = 0$, $\theta = \pi$ of the sphere we observe that the operator Δ is invariant under rotations of the coordinate system; for $r = 1$, Δ becomes Δ^*, hence Δ^* is also invariant under coordinate rotations. The singularities of equation (36) at the points $\theta = 0$, $\theta = \pi$ are a consequence of the particular choice of the coordinate system. It is therefore natural to require that the poles of the sphere be regular points of the function Y. (In other words, Y as a function of position on the sphere should satisfy the same regularity conditions everywhere.)

The eigenvalues λ and associated eigenfunctions Y are most easily determined by investigating those solutions $u = U_n$ of $\Delta u = 0$ which are integral and rational in x, y, z and homogeneous of degree n; this investigation is analogous to §8. If we write these solutions in polar coordinates in the form $U_n = r^n Y_n(\theta, \varphi)$, we see that the Y_n are solutions of equation (36). The eigenvalues corresponding to the $2n + 1$ functions $Y_n(\theta, \varphi)$ are found to be

$$\lambda = n(n + 1).$$

It will be shown in Ch. VII, §5 that the functions Y_n defined in this manner constitute the totality of eigenfunctions of our problem.

Furthermore, it will be possible to give a proof for the completeness and expansion theorems which is similar to that used for the Sturm-Liouville functions. From the expansion theorem it follows that by superposition of solutions: $u = \sum_{n=0}^{\infty} a_n r^n Y_n$, we can find a solution of $\Delta u = 0$ which assumes prescribed values on the surface of the sphere.

Not only the function $u = r^n Y_n$ but also $u = r^{-(n+1)} Y_n$, which has a singularity at zero, is a solution of $\Delta u = 0$. Thus, by superposition of solutions of the form $r^n Y_n$ and $r^{-(n+1)} Y_n$, one can find a solution of $\Delta u = 0$ which assumes prescribed values on two concentric spheres and is regular in the intermediate shell.

If in particular we consider those spherical harmonics which depend only on θ and not on φ, (i.e., if we assume $Y_\varphi = 0$) the differential equation takes the form

$$\frac{1}{\sin\theta}(Y_\theta \sin\theta)_\theta + \lambda Y = 0;$$

performing the transformation $x = \cos\theta$, this equation becomes the equation for the Legendre polynomials (cf. eq. (20), Ch. II). The Legendre polynomials $P_n(\cos\theta)$ are a special case of spherical harmonics.

We obtain a generalization of the spherical harmonics if we consider an arbitrary domain G on the surface of the sphere and attempt to solve the differential equation (36)

$$\Delta^* Y + \lambda Y = 0$$

for a function $Y(\theta, \varphi)$, regular in G, which satisfies homogeneous boundary conditions on the boundary of G; for example, the function may be required to vanish there. Eigenfunctions $Y_1, Y_2, \cdots$ belonging to this domain are called spherical surface harmonics.[1] It follows from the above calculations that, if α and λ satisfy the relation

$$\alpha(\alpha + 1) = \lambda,$$

the function $r^\alpha Y(\theta, \varphi) = u(x, y, z)$ is a solution of the differential equation $\Delta u = 0$ which is continuous, except possibly at the origin, in the cone with G as base and the center of the sphere as vertex.

The equation $\Delta^* Y + \lambda Y = 0$ for spherical surface harmonics is a special case of the general differential equation

$$\Delta^* Y + \lambda Y = \frac{1}{\sqrt{eg - f^2}}\left(\frac{\partial}{\partial y}\frac{e\frac{\partial Y}{\partial y} - f\frac{\partial Y}{\partial x}}{\sqrt{eg - f^2}} + \frac{\partial}{\partial x}\frac{g\frac{\partial Y}{\partial x} - f\frac{\partial Y}{\partial y}}{\sqrt{eg - f^2}}\right) + \lambda Y = 0$$

corresponding to an arbitrary curved surface with the line element $ds^2 = e\,dx^2 + 2f\,dx\,dy + g\,dy^2$. In Ch. IV, §8, 2 we noted that this equation is invariant. It can be considered as the equation of vibration of a "curved membrane" lying on the surface. In the case of a sphere it becomes equation (36) if polar coordinates are introduced.

[1] Cf. W. Thomson and P. G. Tait, Treatise on Natural Philosophy, Vol. I, pp. 171–218, Cambridge University Press, Cambridge, 1886.

2. Cylindrical Domain. A further example of a region for which the boundary value problem of potential theory may be solved explicitly is given by a cylinder erected over the domain G of the x, y-plane which is bounded by the planes $z = 0$ and $z = \pi$. We assume that the boundary values are identically zero on the vertical surface of the cylinder and that they are given on the plane bases by arbitrary twice continuously differentiable functions which vanish on the boundaries Γ. We now seek solutions of $\Delta u = 0$ of the form $u = f(z)v(x, y)$ and obtain as above the conditions $f''/f = -\Delta v/v = \lambda$, i.e., differential equations for f and for v in which λ is to be determined in such a way that an eigenfunction $v(x, y)$ which vanishes on Γ exists. If $v_1, v_2, \cdots$ are all the eigenfunctions, with corresponding eigenvalues $\lambda_1, \lambda_2, \cdots$, then, by the expansion theorem, we can determine the constants a_n, b_n in such a way that the infinite series $\sum_{n=1}^{\infty} (a_n e^{\sqrt{\lambda_n} z} + b_n e^{-\sqrt{\lambda_n} z}) v_n(x, y)$ takes on the given boundary values for $z = 0$ and for $z = \pi$. Hence this series is the solution of our boundary value problem, provided that it converges uniformly together with the series obtained from it by repeated termwise differentiation with respect to any of the variables x, y, z.

3. The Lamé Problem. Essentially, the most general case in which separation of variables helps to reduce the boundary value problem of potential theory to an eigenvalue problem for functions of a single variable is the case of a *confocal rectangular parallelepiped.* By the latter we mean a domain bounded by a portion of each of two ellipsoids, a portion of each of two hyperboloids of one sheet and a portion of each of two hyperboloids of two sheets, all belonging to the same confocal family

$$\frac{x^2}{s - e_1} + \frac{y^2}{s - e_2} + \frac{z^2}{s - e_3} = 1$$

(cf. Ch. IV, §8, 3). *Almost all the boundary value problems which are usually treated explicitly may be considered as special or limiting cases of this "Lamé" problem.* If we introduce elliptic coordinates $\rho = f(u)$, $\sigma = g(v)$, $\tau = h(w)$ in accordance with the notation of Chapter IV, the potential equation $\Delta T = 0$ becomes

$$\Delta T = \frac{[g(v) - h(w)]T_{uu} + [f(u) - h(w)]T_{vv} + [f(u) - g(v)]T_{ww}}{[g(v) - h(w)][f(u) - h(w)][f(u) - g(v)]} = 0.$$

Evidently, solutions of the form

$$T = U(u)V(v)W(w)$$

can be obtained if we can find two constants λ, μ which satisfy the following three ordinary differential equations

$$U'' + [\lambda f(u) + \mu]U = 0, \tag{37}$$

$$V'' - [\lambda g(v) + \mu]V = 0, \tag{38}$$

$$W'' + [\lambda h(w) + \mu]W = 0, \tag{39}$$

where the variables u, v, w lie in the intervals

$$u_2 \leq u \leq u_1, \qquad v_2 \leq v \leq v_1, \qquad w_2 \leq w \leq w_1$$

determined by the conditions

$$\rho_2 \leq f(u) \leq \rho_1, \qquad \sigma_2 \leq g(v) \leq \sigma_1, \qquad \tau_2 \leq h(w) \leq \tau_1.$$

Thus the parallelepiped is given by conditions of the form $\rho_1 \geq \rho \geq \rho_2 \geq \sigma_1 \geq \sigma \geq \sigma_2 \geq \tau_1 \geq \tau \geq \tau_2$.

Using the coordinates ρ, σ, τ instead of u, v, w and denoting the independent variable by s and the dependent variable by Y, we can write equations (37), (38), and (39) in the form

$$\varphi(s)\frac{d^2Y}{ds^2} + \frac{\varphi'(s)}{2}\frac{dY}{ds} + (\lambda s + \mu)Y = 0,$$

where we have set

$$4(s - e_1)(s - e_2)(s - e_3) = \varphi(s).$$

The solutions of this equation (the *Lamé equation*) are functions which depend on the choice of the constants λ, μ, and in general cannot be expressed in terms of the elementary transcendental functions. They are called *Lamé functions* and have been extensively investigated, even though relatively few means for their numerical calculation have been developed. Here we state only the associated eigenvalue problem. Clearly the boundary value problem of potential theory can be solved for a confocal parallelepiped if it can be solved for the special case in which the given boundary values vanish on five of the six faces. Then the solution of the general boundary value problem is the sum of six such special solutions. Suppose, for example, that for $\tau = \tau_2$ nonzero boundary values have been as-

signed. We seek the solutions U, V, W of the Lamé equations (37), (38), (39) for which the relations $U(u_1) = U(u_2) = V(v_1) = V(v_2) = W(w_1) = 0$ hold but place no condition on $W(w_2)$. The product

$$T = U(u)V(v)W(w)$$

will then be a solution of $\Delta T = 0$ which vanishes for $\rho = \rho_2$, $\rho = \rho_1$, $\sigma = \sigma_2$, $\sigma = \sigma_1$, $\tau = \tau_1$. However, as we shall see, the given conditions cannot be satisfied for an arbitrary choice of λ, μ. Thus, we have a new eigenvalue problem, a so-called *two-parameter eigenvalue problem*, in which we must determine pairs of associated eigenvalues λ, μ for which equations (37) and (38) have solutions that vanish for $u = u_1$, $u = u_2$ and $v = v_1$, $v = v_2$, respectively.

The situation in this eigenvalue problem is similar to that in the ordinary one-parameter problem: *There exist infinitely many pairs of eigenvalues* λ_i, μ_i *and corresponding solutions* U_i, V_i *of the eigenvalue problem. Every function of* u, v *which is continuous in the rectangle* $u_2 \leq u \leq u_1$, $v_2 \leq v \leq v_1$ *together with its derivatives up to the second order, and which vanishes on the boundary of the rectangle, can be expanded in an absolutely and uniformly convergent series of the form*

$$\sum_{i=1}^{\infty} c_i U_i(u) V_i(v)$$

where the summation is over all Lamé products $U_i(u)V_i(v)$ *which correspond to the eigenvalue pairs. Furthermore, Lamé products which correspond to different eigenvalue pairs satisfy the orthogonality relation*

$$\int_{u_2}^{u_1} \int_{v_2}^{v_1} [f(u) - g(v)] U_i(u) V_i(v) U_k(u) V_k(v)\, dv\, du = 0.$$

To solve the boundary value problem, the next step is to associate to each pair of eigenfunctions U_i, V_i a function $W_i(w)$ which satisfies equation (39) for $\lambda = \lambda_i$ and $\mu = \mu_i$ and vanishes for $w = w_1$. (The existence of one such solution follows from general existence theorems in the theory of differential equations.) The $W_i(w)$ do not vanish for $w = w_2$, for otherwise $T = UVW$ would be a nonvanishing solution of $\Delta T = 0$ with vanishing boundary values, which contradicts elementary facts of potential theory.

The boundary values assigned on $w = w_2$ can then be expanded in a series of the form

$$\sum_{i=1}^{\infty} a_i W_i(w_2) U_i(u) V_i(v),$$

and the series

$$\sum_{i=1}^{\infty} a_i U_i(u)V_i(v)W_i(w)$$

represents the desired solution of the boundary value problem of potential theory for our parallelepiped. The above formulation of the expansion theorem apparently implies that the prescribed boundary values of T should vanish on all the edges of the parallelepiped. Further investigation would show, however, that this restriction need not actually be imposed.

Our two-parameter eigenvalue problem is easily reduced to a one-parameter problem for a partial differential equation. Let us set $Z(u, v) = U(u)V(v)$, where $U(u)$ is a solution of equation (37) and $V(v)$ a solution of equation (38). If we multiply the first of these equations by V, the second by U, and add, we obtain the partial differential equation

$$Z_{uu} + Z_{vv} + \lambda[f(u) - g(v)]Z = 0 \tag{40}$$

for the function $Z(u, v)$. This differential equation might have been obtained directly from $\Delta T = 0$ by writing $T = Z(u, v)W(w)$. Evidently, the eigenvalue $\lambda = \lambda_i$ and the associated eigenfunction $Z_i = U_i(u)V_i(v)$ solve the eigenvalue problem of this differential equation for the rectangle G: $u_2 \leq u \leq u_1$, $v_2 \leq v \leq v_1$ under the boundary condition $Z = 0$. Since equation (40) has the form $\Delta Z + \lambda\rho Z = 0$, where the function $\rho = f(u) - g(v)$ is positive throughout the rectangle F, this eigenvalue problem is of the type considered above, and we have analogous questions of existence of eigenfunctions and validity of an eigenfunction expansion theorem. These questions will be discussed later.[1] At present we postulate that, for the rectangle G, infinitely many eigenvalues $\lambda_1, \lambda_2, \cdots$ exist together with associated eigenfunctions $Z_1, Z_2, \cdots$ which vanish at the boundary; arbitrary functions can be expanded in terms of the λ_i and Z_i as described above. Now we shall show that all the eigenfunctions Z_i are Lamé products $U(u)V(v)$, or the sum of at most a finite number of Lamé products belonging to the same eigenvalue λ.

Proof: let $\lambda_1, \lambda_2, \cdots$ and $Z_1, Z_2, \cdots$ denote the complete system of eigenvalues and corresponding eigenfunctions for (40). Corre-

[1] See §§14, 15.

sponding to any eigenvalue λ_h, we consider the ordinary differential equation

$$\frac{d^2X}{du^2} + [\lambda_h f(u) + \mu]X = 0$$

with the boundary condition $X = 0$ for $u = u_1$ and $u = u_2$. We denote the infinitely many associated eigenvalues and the normalized eigenfunctions by $\mu_1, \mu_2, \cdots$ and $X_1, X_2, \cdots$, respectively. Given any function which vanishes for $u = u_1$ and $u = u_2$ and is continuous with its derivatives up to the second order in the interval $u_2 \leq u \leq u_1$, we can expand it in terms of these eigenfunctions. This is true in particular for the function $Z(u, v)$ which depends on an additional parameter v. We write the expansion in the form

$$Z(u, v) = \sum_{n=1}^{\infty} Y_n(v)X_n(u),$$

where

$$Y_n(v) = \int_{u_2}^{u_1} Z(u, v)X_n(u)\, du$$

and differentiate Y_n twice with respect to v. Using integration by parts, we obtain

$$\begin{aligned}
\frac{d^2Y_n}{dv^2} &= \int_{u_2}^{u_1} Z_{vv}(u, v)X_n(u)\, du \\
&= \int_{u_2}^{u_1} (-Z_{uu} - \lambda_h[f(u) - g(v)]Z)X_n\, du \\
&= \int_{u_2}^{u_1} Z\left(-\frac{d^2X_n}{du^2} - \lambda_h[f(u) - g(v)]X_n\right) du \\
&= (\mu_n + \lambda_h g(v))\, Y_n .
\end{aligned}$$

This means that Y_n is an eigenfunction for equation (38) for the domain $v_2 \leq v \leq v_1$ with the given boundary condition. In other words, the numbers λ_h, μ_n with the associated functions $X_n(u)$, $Y_n(v)$ constitute a solution of our two-parameter eigenvalue problem—provided the function $Y_n(v)$ does not vanish identically. As we have already seen, each product X_nY_n is an eigenfunction of (40) which corresponds to the eigenvalue λ_h. However, the eigenvalues of this differential equation can have only a finite multiplicity, as will be

shown in the next chapter. Hence only a finite number k of the functions X_n, Y_n can be linearly independent. We may suppose further that no X_n or Y_n vanishes identically for otherwise we could simply omit the term in question. It follows that any $k+1$ products X_nY_n satisfy a linear relation

$$\sum_{\nu=1}^{k+1} c_\nu X_{n_\nu} Y_{n_\nu} = 0.$$

If we assign values to the variables v for which all Y_{n_ν} are different from zero we obtain a linear equation between the X_{n_ν}. This, however, is impossible since eigenfunctions which correspond to different eigenvalues μ are linearly independent. Hence, the representation $Z = \sum X_nY_n$ can have at most k terms, which is what we wanted to show.

We can now expand an arbitrary function, subject to the indicated restrictions, in terms of the eigenfunctions Z_i and obtain the following result: Any function which is continuous, together with its derivatives up to the second order, in the rectangle $u_2 \leq u \leq u_1$, $v_2 \leq v \leq v_1$ and vanishes on the boundary of the rectangle, can be expanded in a series of Lamé products.

§10. *Problems of the Sturm-Liouville Type. Singular Boundary Points*

The separation of variables sometimes leads to eigenvalue problems for differential equations of the Sturm-Liouville type,

$$(pu')' - qu + \lambda\rho u = 0,$$

for which singularities occur at end points of the fundamental domain. The coefficient p, for example, may vanish at an end point. The nature of the problem is such that certain conditions are imposed at the singular end points; e.g. the solution must be continuous or bounded or become infinite of an order less than that prescribed. Conditions of this type replace the homogeneous boundary conditions considered in §3.

1. Bessel Functions. Consider for example the Bessel equation (cf. §5, 5)

$$(xu')' - \frac{n^2}{x}u + \lambda xu = 0, \tag{41}$$

which occurs very frequently in mathematical physics. For this equation the assumption of §3, 3, that $p > 0$ throughout the entire fundamental domain $0 \leq x \leq 1$, is not valid because $p(0) = 0$. In other words, the point $x = 0$ is a singular point for the Bessel equation and the requirement that a solution remain finite at this point represents a boundary condition. In this case we have the problem of finding a solution which remains finite for $x = 0$ and, e.g., vanishes for $x = 1$. The eigenfunctions are Bessel functions $J_n(\sqrt{\lambda}x)$, where the eigenvalues $\lambda = \lambda_{n,m}$ are determined as the roots of a transcendental equation by the boundary condition at $x = 1$.

The associated orthogonal functions $z = \sqrt{x}J_n(\sqrt{\lambda}x)$ are characterized by means of the differential equation

$$z'' - \frac{n^2 - \frac{1}{4}}{x^2} z + \lambda z = 0 \tag{42}$$

which is obtained directly from the Bessel equation. (This is an example of the transformation given in a general form on page 292.) For the function $\zeta = z/x = J_n(\sqrt{\lambda}x)/\sqrt{x}$ we obtain the differential equation

$$(x^2\zeta')' - (n^2 - \tfrac{1}{4})\zeta + \lambda x^2\zeta = 0. \tag{43}$$

2. Legendre Functions of Arbitrary Order. A similar type of problem is presented by the Sturm-Liouville equation

$$[(1 - x^2)u']' + \lambda u = 0 \tag{44}$$

with the boundary conditions that u remain finite for $x = +1$ and for $x = -1$, i.e. at the two singularities of the equation. The fundamental domain is $-1 \leq x \leq +1$. We know from Ch. II, §8 that the numbers $\lambda = n(n + 1)$ are eigenvalues and the Legendre functions $P_n(x)$ eigenfunctions.

It is easy to show that *the Legendre polynomials are the only solutions of this eigenvalue problem.* For example, this can be deduced from the fact that the Legendre functions form a complete orthogonal system as shown in Ch. II, §8. We shall give an alternative proof, which is independent of this fact: We note that, whenever a function $u = f(x)$ satisfies equation (44), then the function $u = f(-x)$ also satisfies this equation. Evidently, the functions $f(x) + f(-x)$ and $f(x) - f(-x)$ are also solutions; one of them is an even, the other an odd function, and at least one does not vanish identically

since by assumption u is not identically equal to zero. Thus we need only show that every even and every odd solution u of (44) which is continuous for $-1 \leq x \leq 1$ is a Legendre polynomial, and that λ must be a number of the form $n(n+1)$. If we write the solution (which is analytic) as a power series: $u = \sum_{\nu=0}^{\infty} a_\nu x^\nu$, equation (44) leads at once to the recursion formula

$$a_\nu = \frac{(\nu-1)(\nu-2) - \lambda}{\nu(\nu-1)} a_{\nu-2}. \tag{45}$$

If u is even, all the a_ν for which ν is odd are zero. If u is odd, the same is true for the a_ν for which ν is even. If $\nu - 2h > 0$, it follows immediately from (45) that

$$a_\nu = \frac{1}{\nu}\left[1 - \frac{\lambda}{(\nu-1)(\nu-2)}\right]\left[1 - \frac{\lambda}{(\nu-3)(\nu-4)}\right] \cdots \left[1 - \frac{\lambda}{(\nu-2h+1)(\nu-2h)}\right] k a_k, \tag{46}$$

where $k = \nu - 2h$. The series for u has a finite number of terms if and only if λ has the form $n(n+1)$. In this case we see immediately that u represents the n-th Legendre polynomial. For all other values of λ we obtain an infinite power series, which by elementary criteria converges for $|x| < 1$. Fix k so large that all the factors of the above product are positive (a_k may be assumed positive). By well-known theorems, the product of the bracketed factors on the right side of (46) converges to a positive limit as $\nu \to \infty$; hence, $a_\nu > c/\nu$ for $\nu > k$, where c is a positive constant. It follows that the absolute value of $\sum_{n=k}^{\nu} a_n x^n$ will be arbitrarily large if $|x|$ is sufficiently close to 1 and ν sufficiently large. This implies that $\lim_{x \to \pm 1} |u(x)| = \infty$ and thus that λ is not an eigenvalue.[1]

We can easily derive other classes of orthogonal systems of eigenfunctions from the differential equation for the Legendre polynomials by a general method. Namely, if we differentiate equation (44) with respect to x, we obtain a differential equation for the function $u'(x)$. As before it follows that a solution which is regular at both end points of the interval exists only for $\lambda = n(n+1)$ and is given

[1] The above discussion is closely related to the Raabe or Gauss convergence criterion; cf. A. Kneser, Zur Theorie der Legendreschen Polynome, Tôhoku Math. J., Vol. 5, 1914, pp. 1–7.

in this case by $P_n'(x)$. The resulting equation for $P_n'(x)$ is not yet self-adjoint; we make it self-adjoint by introducing the function $P_n'(x)\sqrt{1-x^2} = z_n$ as the unknown; then the new equation takes the form

$$[(1-x^2)z']' - \frac{z}{1-x^2} + \lambda z = 0.$$

The associated eigenvalues are $\lambda = n(n+1)$ $(n = 1, 2, 3, \cdots)$ with the eigenfunctions

$$z_n = \sqrt{1-x^2}\, P_n'(x).$$

The functions $z_n = P_{n,1}(x)$ are called *associated Legendre functions of first order*. (The functions $P_n(x) = P_{n,0}(x)$ will occasionally be referred to as Legendre functions of zero-th order.) The Legendre functions $P_{n,1}$ satisfy the orthogonality relation

$$\int_{-1}^{1} P_{n,1}P_{m,1}\, dx = 0 \quad \text{for} \quad n \neq m.$$

In a similar way, by differentiating (44) h times we obtain for the function

$$(1-x^2)^{h/2}\frac{d^h}{dx^h}P_n(x) = P_{n,h}(x)$$

the differential equation

$$[(1-x^2)z']' - \frac{h^2 z}{1-x^2} + \lambda z = 0, \tag{47}$$

with eigenvalues $\lambda = n(n+1)$ $(n = h, h+1, \cdots)$ and associated eigenfunctions $P_{n,h}(x)$, which also are mutually orthogonal. These functions are called associated *Legendre functions of h-th order*. They can be normalized with the aid of the easily verified equation

$$\int_{-1}^{1} P_{n,h}^2\, dx = \frac{2}{2n+1}\frac{(n+h)!}{(n-h)!}.$$

We can prove that we have obtained all eigenvalues and eigenfunctions of (47) by the method used for Legendre polynomials.

3. Jacobi and Tchebycheff Polynomials. A generalization of the Legendre polynomials is given by *Jacobi's polynomials* of Ch. II, §9.

The differential equation for these polynomials may be written in the following Sturm-Liouville form:

$$[(1 - x)^{p-q+1}x^q u']' + \lambda(1 - x)^{p-q}x^{q-1}u = 0.$$

The n-th Jacobi polynomial corresponds to the eigenvalue $\lambda = n(p + n)$ with the boundary conditions that the solution remain finite for $x = 0$ and $x = 1$. As above, there are two ways to show that Jacobi's polynomials are the only solutions of this eigenvalue problem.

A further example is offered by the Tchebycheff polynomials, which correspond to the Sturm-Liouville equation

$$(\sqrt{1 - x^2}\, u')' + \frac{\lambda}{\sqrt{1 - x^2}}u = 0$$

with the boundary conditions that the solution be regular at $x = \pm 1$. The eigenvalue corresponding to the Tchebycheff polynomial $T_n(x)$ is $\lambda = n^2$ and, as above, these λ and T_n exhaust all the eigenvalues and eigenfunctions.

4. Hermite and Laguerre Polynomials. The Hermite polynomials $u = H_n(x)$ and the corresponding orthogonal functions $v = H_n e^{-x^2/2}$ are characterized as the solutions of the eigenvalue problems (cf. Ch. II, §9, 4)

$$(e^{-x^2}u')' + \lambda e^{-x^2}u = 0 \tag{48}$$

and

$$v'' + (1 - x^2)v + \lambda v = 0, \tag{49}$$

respectively, with eigenvalues $\lambda = 0, 2, 4, \cdots$. The fundamental domain is the entire line $-\infty < x < +\infty$, and the boundary condition for (48) is: the eigenfunction u should not become infinite at $x = \pm\infty$ of an order higher than a finite power of x. To show that the Hermite eigenvalue problem has no other solutions we write equation (48) in the form $u'' - 2xu' + \lambda u = 0$ and u as a power series $u = \sum_{n=0}^{\infty} a_n x^n$. We can assume that u is either an even or an odd function (see the discussion of eq. (44)), and hence that only odd or only even powers of x appear in the power series. The differential equation implies the recursion formula

$$\frac{a_{n+2}}{a_n} = \frac{2n - \lambda}{(n + 1)(n + 2)}$$

for the nonvanishing coefficients. Hence, either the series breaks off—in case $\lambda = 2n$ is a non-negative even integer—and thus represents the Hermite polynomial H_n, or the series has infinitely many nonvanishing coefficients and converges for all values of x. As soon as $2n - \lambda$ becomes positive, all the coefficients a_n which occur have the same sign. In the second case, terms $a_n x^n$ with n arbitrarily large occur; therefore u becomes infinite at $x = +\infty$ of an order greater than any finite power of x. Thus, u cannot be an eigenfunction for the problem and the Hermite polynomials are the only solutions of the eigenvalue problem.

The *Laguerre polynomials* will be treated in more detail, since an application will be given in §12, 4. Here the fundamental domain is the positive real axis $0 \leq x < \infty$, and according to Ch. II, §9 the eigenvalue equation, satisfied by the Laguerre polynomials $u = L_n(x)$ for the eigenvalue $\lambda = n$ (n a positive integer), is of the form

$$xu'' + (1 - x)u' + \lambda u = 0 \tag{50}$$

or, in self-adjoint form, it is

$$(xe^{-x}u')' + \lambda e^{-x}u = 0,$$

where the boundary conditions are: u remains finite at $x = 0$ and, as $x \to \infty$, u does not become infinite of an order higher than a positive power of x. For the associated orthogonal functions

$$v = \omega_n = e^{-x/2}L_n$$

we find the Sturm-Liouville eigenvalue equation

$$(xv')' + \left(\frac{1}{2} - \frac{x}{4}\right)v + \lambda v = 0,$$

where we require regularity at $x = 0$ as the boundary condition. Finally, we remark that the functions

$$w = S_n = x^{-1/2}\omega_n$$

which occur later (§12, 4) satisfy the self-adjoint eigenvalue equation

$$(x^2w')' - \frac{x^2 - 2x - 1}{4}w + \lambda xw = 0$$

where the solution is required to vanish at $x = 0$. The corresponding eigenvalues are, throughout, the positive integers $\lambda = n$.

Here, as in the case of Legendre functions in subsection 2, processes of differentiation and multiplication by suitable factors lead to Laguerre functions of higher order which satisfy similar differential equations. By differentiating (50) m times, we find that the functions

$$u = L_n^m(x) = \frac{d^m}{dx^m} L_n(x)$$

satisfy the differential equation

$$xu'' + (m + 1 - x)u' + (\lambda - m)u = 0, \tag{51}$$

which can be written in the following self-adjoint form:

$$(x^{m+1}e^{-x}u')' + x^m e^{-x}(\lambda - m)u = 0.$$

The associated orthogonal functions

$$v = \omega_n^m = x^{m/2}e^{-x/2}L_n^m$$

satisfy the Sturm-Liouville equation

$$(xv')' + \left(\frac{1 - m}{2} - \frac{x}{4} - \frac{m^2}{4x}\right)v + \lambda v = 0, \tag{51a}$$

and the functions

$$w = S_n^m = x^{-1/2}\omega_n^m$$

satisfy the eigenvalue equation

$$(x^2w')' - \frac{x^2 + 2(m - 1)x + m^2 - 1}{4}w + \lambda xw = 0 \tag{51b}$$

for the corresponding eigenvalues $\lambda = n$, where n is an integer greater than or equal to m and the boundary conditions are evident.

In order to show that our differential equation has no other eigenvalues or eigenfunctions, we set $u = \sum_{\nu=0}^{\infty} a_\nu x^\nu$ in equation (51) and find (using recursion relations) the formula

$$a_\nu = \frac{a_0}{\nu!} \frac{(m - \lambda) \cdots (m - \lambda + \nu - 1)}{(m + 1) \cdots (m + \nu)}$$

for the coefficients. We see that for arbitrary fixed λ the coefficients of this series have the same sign from a certain ν on, and that the series converges for all values of x. Thus it does, in fact, represent a

solution of (51) which is regular for $0 \leq x < \infty$. In the case of positive integral $\lambda = n$, with $n > m$, the series breaks off after a finite number of terms and thus represents a polynomial. For every other λ it is easy to obtain the estimate

$$|a_\nu| > \frac{c}{\nu! \nu^r},$$

where c is a suitable constant and r a suitable positive integral exponent. But this implies that for $x \to \infty$ the solution becomes infinite of at least the order of e^x/x^r. Hence this solution cannot be an eigenfunction for the problem, and our assertion is proved.

§11. *The Asymptotic Behavior of the Solutions of Sturm-Liouville Equations*

Simple asymptotic evaluations of the solutions of Sturm-Liouville differential equations can be obtained for large values of the parameter or independent variable provided the coefficients satisfy certain general conditions.

1. Boundedness of the Solution as the Independent Variable tends to Infinity. Let us write the differential equation in the form $u'' + \mu(x)u = 0$ (cf. §3, eq. (19a)). We assume that for $x \to \infty$, $\mu(x)$ approaches a positive limit which, without loss of generality, can be taken equal to 1. Then, setting $\mu = 1 + \rho$, we can base our discussion on the differential equation

$$u'' + u + \rho u = 0. \tag{52}$$

We shall, however, replace the assumption $\rho \to 0$ by the more stringent one

$$|\rho| < \frac{\alpha}{x}, \qquad |\rho'| < \frac{\alpha}{x^2}, \tag{53}$$

where α is a positive constant. With this assumption, we assert that every solution of equation (52) is bounded for $x \to \infty$. This is to be expected since, for large x, (52) approaches the equation $u'' + u = 0$, which has only bounded solutions.

To prove our assertion, we multiply (52) by u', integrate from a positive lower limit a (to be suitably determined later) to x, and obtain

$$u'^2\Big|_a^x + u^2\Big|_a^x = -2\int_a^x \rho u u'\,dx = -\rho u^2\Big|_a^x + \int_a^x \rho' u^2\,dx. \tag{54}$$

From this it follows immediately that

$$u^2(x) \leq u'^2(x) + u^2(x) \leq C(a) + |\rho(x)|\, u^2(x) + \int_a^x |\rho'|\, u^2\, dx,$$

where $C(a)$ denotes an expression which depends only on the lower limit a. Let $M = M(x)$ be the maximum of the function $u(\xi)$ in the interval $a \leq \xi \leq x$, assumed at the point $\bar{\xi}$; then from this inequality and from (53) it follows that

$$M^2 \leq C(a) + \frac{M^2\alpha}{\xi} + M^2\alpha\left(\frac{1}{a} - \frac{1}{\xi}\right)$$

and thus that

$$M^2\left(1 - \frac{\alpha}{a}\right) \leq C(a).$$

Now if we take $a \geq 2\alpha$ we obtain $2C(a)$ as a bound for M^2, independent of x. This proves the assertion.

2. A Sharper Result. (Bessel Functions.) We consider the equation $u'' + u + \rho u = 0$ and assume that $\rho(x)$ vanishes at infinity of an order higher than the first (this assumption is more stringent than that made in subsection 1); i.e. we assume for example[1]

$$\rho(x) = O\left(\frac{1}{x^2}\right) \tag{55}$$

We now have a closer agreement between the differential equation and the equation $u'' + u = 0$, which implies not only that the solutions are bounded but also that they approach trigonometric functions asymptotically.

We proceed by determining functions $\alpha(x)$ and $\delta(x)$ with the derivatives α' and δ', related to u and u' by

$$u = \alpha \sin(x + \delta), \qquad u' = \alpha \cos(x + \delta)$$

(α cannot vanish at any point since u and u' would then vanish simultaneously at some point and thus u would vanish identically in virtue of equation (52)). We can calculate u'' and u' in two ways:

[1] Here we use the customary notation, according to which $O(f(x))$ denotes a function $g(x)$ for which the quotient $|g(x)/f(x)|$ remains bounded as the argument increases.

we obtain

$$u'' = \alpha' \cos(x + \delta) - \alpha(\delta' + 1) \sin(x + \delta)$$
$$= -(1 + \rho)\, \alpha \sin(x + \delta),$$

$$\tan(x + \delta) = \frac{\alpha'}{\alpha(\delta' - \rho)};$$

$$u' = \alpha \cos(x + \delta) = \alpha' \sin(x + \delta) + \alpha(\delta' + 1) \cos(x + \delta),$$

$$\tan(x + \delta) = -\frac{\alpha\delta'}{\alpha'},$$

$$\tan^2(x + \delta) = -\frac{\delta'}{\delta' - \rho},$$

$$\delta' = \rho \sin^2(x + \delta), \tag{56}$$

$$\frac{\alpha'}{\alpha} = \frac{-\delta'}{\tan(x + \delta)} = -\rho \sin(x + \delta) \cos(x + \delta). \tag{57}$$

We see that α and δ approach definite limits as $x \to \infty$. In fact, $\delta(x) = \delta(\beta) - \int_x^\beta \delta'(\xi)\, d\xi$; if we let β increase beyond all bounds the integral on the right side converges by (55) and (56), since the integrand approaches zero like $1/x^2$. Thus $\lim_{\beta \to \infty} \delta(\beta) = \delta_\infty$ exists, and the above representation shows moreover that

$$\delta(x) = \delta_\infty + O\left(\frac{1}{x}\right).$$

Correspondingly, formula (57) for $\alpha'/\alpha = d \log \alpha/dx$ leads to the relation

$$\alpha(x) = \alpha_\infty\left(1 + O\left(\frac{1}{x}\right)\right),$$

in which $\alpha_\infty \neq 0$. Thus, for every solution u we have obtained the asymptotic representation

$$u = \alpha \sin(x + \delta) = \alpha_\infty \sin(x + \delta_\infty) + O\left(\frac{1}{x}\right).$$

This result may be immediately applied to the differential equation

$$u'' + \left(1 - \frac{m^2 - \frac{1}{4}}{x^2}\right)u = 0,$$

whose solutions, according to page 325, are connected with the solutions $y_m(x)$ of the Bessel equation by the relations

$$u = y_m\sqrt{x}.$$

Thus all solutions $y_m(x)$ of the Bessel equation satisfy asymptotic formulas of the form

$$y_m(x) = \frac{\alpha_\infty}{\sqrt{x}} \cos(x + \delta_\infty) + O\left(\frac{1}{x^{3/2}}\right).$$

The constants α_∞ and δ_∞ for the Bessel function $J_m(x)$ will be determined from other considerations (cf. Ch. VII, §6, 2). It will be found that

$$\alpha_\infty = \sqrt{\frac{2}{\pi}}, \qquad \delta_\infty = -\frac{m\pi}{2} - \frac{\pi}{4}.$$

3. Boundedness as the Parameter Increases. Considerations similar to those of subsection 1 are used to prove the following theorem: *The absolute values of the solutions of the Sturm-Liouville equation (with continuous r)*

$$u'' - ru + \lambda u = 0 \tag{58}$$

in the interval $0 \leq x \leq 1$ remain less than some bound independent of λ and x provided the solutions are normalized by the condition

$$\int_0^1 u^2\,dx = 1$$

and satisfy the boundary conditions $u(0) = u(1) = 0$.

It suffices to prove the theorem for large positive λ; we again multiply the equation by u' and integrate from 0 to x, obtaining

$$u'^2(x) + \lambda u^2(x) - 2\int_0^x ruu'\,dx = u'^2(0) + \lambda u^2(0). \tag{59}$$

To evaluate the right side we integrate this equation between the limits 0 and 1, obtaining

$$u'^2(0) + \lambda u^2(0) = \int_0^1 u'^2\,dx + \lambda - 2\int_0^1 dx \int_0^x ruu'\,dt. \tag{60}$$

Inserting this value in (59) and estimating the resulting integrals by means of Schwarz's inequality, we have

$$(61) \quad \lambda u^2 \leq u'^2 + \lambda u^2 \leq \lambda + \int_0^1 u'^2\,dx + C_1 \sqrt{\int_0^1 u'^2\,dx}\sqrt{\int_0^1 u^2\,dx},$$

where C_1 denotes a positive constant independent of x and λ. From the equation

$$\int_0^1 u'^2\,dx + \int_0^1 ru^2\,dx = \lambda,$$

obtained in the familiar manner by multiplying (58) by u and transforming the resulting expression by Green's formula, we have

$$\int_0^1 u'^2\,dx \leq \lambda + C_2 \int_0^1 u^2\,dx.$$

Inserting this in (61) we obtain the inequality

$$\lambda u^2(x) \leq 2\lambda + C_3\sqrt{\lambda} + C_4,$$

where C_2, C_3, C_4 again are positive constants independent of x and λ. This implies

$$u^2(x) \leq 2 + \frac{C_3}{\sqrt{\lambda}} + \frac{C_4}{\lambda},$$

and our assertion is proved.

Finally, we remark that both our result and our method of proof remain valid if no boundary conditions are imposed. However, functions of more than one variable do not have the analogous boundedness property.[1]

4. Asymptotic Representation of the Solutions. Having shown that the solutions are bounded, we shall now prove the following theorem: If u is any normalized solution of $u'' - ru + \lambda u = 0$ in the interval $0 \leq x \leq 1$, with $\lambda > 0$, then there exists a solution v of $v'' + \lambda v = 0$ such that

$$u = v + O\left(\frac{1}{\sqrt{\lambda}}\right).$$

[1] A simple counter-example is provided by the normalized eigenfunctions $(\sqrt{2}/J_0(k_{0,m}r)/J_0'(k_{0,m})$ of the differential equation $\Delta u + \lambda u = 0$, which vanish on the boundary of the unit circle. (Cf. W. Sternberg, Über die asymptotische Integration partieller Differentialgleichungen II, Math. Ann., Vol. 86, particularly pp. 292–295.)

For large values of λ this formula furnishes an asymptotic representation of the solutions u in terms of the trigonometric functions v. To prove the theorem, consider the solution of $v'' + \lambda v = 0$ for which $u(0) = v(0)$, $u'(0) = v'(0)$. The function $u - v = w$ is then a solution of the equation

$$w'' + \lambda w = ru.$$

If we multiply this equation by $2w'$ and integrate from 0 to x, taking into account the fact that $w(0) = w'(0) = 0$, we obtain

$$w'^2(x) + \lambda w^2(x) = 2\int_0^x ruw'\,dx. \tag{62}$$

Let M denote the maximum of $|w(x)|$ and M' the maximum of $|w'(x)|$ in the interval $0 \leq x \leq 1$. Schwarz's inequality applied to (62) yields, because of $\lambda > 0$,

$$M'^2 \leq M'C, \qquad M' \leq C,$$

where C is some positive constant independent of λ and x. It follows from equation (62) that

$$\lambda M^2 \leq C^2$$

and hence

$$M \leq \frac{C}{\sqrt{\lambda}},$$

which completes the proof.

5. Asymptotic Representation of Sturm-Liouville Eigenfunctions. The problem of representing solutions of the equation $(pu')' - qu + \lambda u = 0$ asymptotically may be formulated somewhat differently if, instead of considering arbitrary solutions, we consider eigenfunctions, say for the interval $0 \leq x \leq \pi$ and boundary conditions $u(0) = u(\pi) = 0$. Suppose that by (20a) we transform the differential equation into

$$z'' - rz + \lambda z = 0, \tag{63}$$

where the new independent variable t ranges over the interval $0 \leq t \leq l$ and r is a continuous function in this interval. We wish to compare the n-th eigenfunction z_n, associated with the eigenvalue λ_n, with the corresponding n-th eigenfunction of the differential equation $v'' + \lambda v = 0$.

If, in equation (10), we replace the function N_i by rz we obtain a useful tool: The solutions of (63) which vanish at $t = 0$ satisfy the "Volterra integral equation" for z

$$z(t) = a \sin \sqrt{\lambda}\, t + \frac{1}{\sqrt{\lambda}} \int_0^t r(\tau) z(\tau) \sin \sqrt{\lambda}(t - \tau)\, d\tau \tag{64}$$

with arbitrary constant a.

According to subsection 3 the functions $z(t)$ satisfying (64) and the boundary condition $z(l) = 0$ remain bounded for all λ; it follows immediately from (64) that a is also bounded.[1] This fact, together with equation (64) and the relation $\int_0^l z^2\, dt = 1$, leads to the precise estimate

$$a = \sqrt{\frac{2}{l}} + O\left(\frac{1}{\sqrt{\lambda}}\right)$$

for a, which in turn implies

$$z(t) - \sqrt{\frac{2}{l}} \sin \sqrt{\lambda} t = O\left(\frac{1}{\sqrt{\lambda}}\right).$$

Since the n-th eigenvalue λ_n of the differential equation becomes infinite as n increases (cf. Ch. VI, §2, 2), we see immediately that the n-th eigenfunction $z_n(t)$ has the asymptotic representation

$$z_n(t) = \sqrt{\frac{2}{l}} \sin \sqrt{\lambda_n}\, t + \frac{1}{\sqrt{\lambda_n}} O(1).$$

Furthermore we have the asymptotic estimate for λ_n (cf. Ch. VI, §2, 3)

$$\lambda_n = n^2 \frac{\pi^2}{l^2} + O(1).$$

Hence $\sqrt{\lambda_n} = n(\pi/l) + O(1/n)$, so that

$$\sin \sqrt{\lambda_n} t = \sin n \frac{\pi}{l} t + O\left(\frac{1}{n}\right).$$

Accordingly, the normalized eigenfunctions of the equation $z'' - rz + \lambda z = 0$ have the asymptotic representation

$$z_n(t) = \sqrt{\frac{2}{l}} \sin n \frac{\pi}{l} t + O\left(\frac{1}{n}\right). \tag{65}$$

[1] The boundedness of $z(t)$ may also be proved directly from the integral representation (64).

Similarly, if we differentiate the integral equation (64) we obtain the corresponding formula

$$z_n'(t) = n\frac{\pi}{l}\sqrt{\frac{2}{l}}\cos n\frac{\pi}{l}t + O(1). \tag{66}$$

In terms of the original equation, these results are expressed by the relations

$$u_n(x) = c_n \frac{\sin\left(n\frac{\pi}{l}\int_0^x \sqrt{\frac{\rho}{p}}\,dx\right)}{\sqrt[4]{p\rho}} + O\left(\frac{1}{n}\right), \tag{67}$$

where the normalizing factor c_n is determined by

$$\frac{1}{c_n^2} = \int_0^\pi \frac{\sin^2\left(n\frac{\pi}{l}\int_0^x \sqrt{\frac{\rho}{p}}\,dx\right)}{\sqrt{p\rho}}\,dx$$

and

$$l = \int_0^\pi \sqrt{\frac{\rho}{p}}\,dx$$

holds. Correspondingly, we have

$$u_n'(x) = c_n\frac{n\pi}{l}\frac{\cos\left(n\frac{\pi}{l}\int_0^x \sqrt{\frac{\rho}{p}}\,dx\right)}{\sqrt[4]{p\rho}}\sqrt{\frac{\rho}{p}} + O(1). \tag{68}$$

Asymptotic expressions for the eigenfunctions and their derivatives in the case of more general homogeneous boundary conditions are derived in the same way. We obtain the expressions

$$z_n(t) = \sqrt{\frac{2}{l}}\cos n\frac{\pi}{l}t + O\left(\frac{1}{n}\right) \tag{69}$$

and

$$z_n'(t) = -\frac{n\pi}{l}\sqrt{\frac{2}{l}}\sin n\frac{\pi}{l}t + O(1), \tag{70}$$

which are valid as long as the coefficient h in the boundary condition $z'(0) - hz(0) = 0$ remains finite.

Incidentally, the Volterra integral equation (64) makes it possible to obtain much more precise expressions for the eigenfunctions and

their derivatives. This may be expected because the Neumann series for such a Volterra integral equation always converges (cf. Ch. III, §10, 8).[1] We can obtain these expressions directly, without referring to the general theory. In (64) we set a equal to 1, i.e. we no longer require that the functions be normalized; under the integral sign in the right side we then substitute for $z(\tau)$ the value given by the integral equation. If we iterate this process and set $v(t) = \sin\sqrt{\lambda}t$, we obtain formula

$$
\begin{aligned}
z(t) = v(t) &+ \frac{1}{\sqrt{\lambda}} \int_0^t d\tau_1 v(\tau_1) r(\tau_1) \sin \sqrt{\lambda}(t - \tau_1) \\
&+ \frac{1}{\lambda} \int_0^t d\tau_1 \int_0^{\tau_1} d\tau_2 v(\tau_2) r(\tau_1) r(\tau_2) \sin \sqrt{\lambda}(t - \tau_1) \\
&\qquad \cdot \sin \sqrt{\lambda}(\tau_1 - \tau_2) \\
&+ \frac{1}{\sqrt{\lambda^3}} \int_0^t d\tau_1 \int_0^{\tau_1} d\tau_2 \int_0^{\tau_2} d\tau_3 v(\tau_3) r(\tau_1) r(\tau_2) r(\tau_3) \\
&\qquad \cdot \sin \sqrt{\lambda}(t - \tau_1) \sin \sqrt{\lambda}(\tau_1 - \tau_2) \sin \sqrt{\lambda}(\tau_2 - \tau_3) \\
&+ \cdots \\
&+ \frac{1}{\sqrt{\lambda^n}} \int_0^t d\tau_1 \cdots \int_0^{\tau_{n-1}} d\tau_n z(\tau_n) r(\tau_1) \cdots r(\tau_n) \\
&\qquad \cdot \sin \sqrt{\lambda}(t - \tau_1) \cdots \sin \sqrt{\lambda}(\tau_{n-1} - \tau_n).
\end{aligned} \tag{71}
$$

This series can be continued indefinitely in decreasing powers of $\sqrt{\lambda}$; the result is an infinite series for $z(t, \lambda)$ in terms of decreasing powers of $\sqrt{\lambda}$ which converges for all $\lambda > 0$. The error which results if we take only the first n terms is of a smaller order of magnitude than $(1/\sqrt{\lambda})^n$.

§12. *Eigenvalue Problems with a Continuous Spectrum*

The eigenvalues of the problems considered previously form a denumerably infinite sequence. However, if the coefficients of the differential equation are singular at the boundary points of the fundamental domain or, in particular, if the fundamental interval is infinite the *spectrum*, or totality of eigenvalues, may behave

[1] Cf. J. Liouville, J. de math. pures et appl., Vols. 1, 2, 1836/37 (see bibliography for Ch. VI), where Volterra's integral equation and the Neumann series occur.

quite differently. In particular, *continuous spectra* may occur, i.e. spectra which contain all the values of a λ-interval; in this case a Fourier integral theorem replaces the eigenfunction expansion theorem.

1. Trigonometric Functions. A simple problem of this kind is given by the eigenvalue equation

$$u'' + \lambda u = 0$$

for the interval $-\infty < x < \infty$ with the boundary condition: u remains bounded at plus and minus infinity. Clearly, any non-negative number λ is an eigenvalue, with the associated eigenfunctions $\sin \sqrt{\lambda}\, x$, $\cos \sqrt{\lambda}\, x$. For this eigenvalue problem the special Fourier integral theorem of Ch. II, §6 takes the place of the expansion theorem.

In general, the possibility of continuous spectra becomes plausible through a limiting process. We start with an eigenvalue problem for a finite interval and then pass to the infinite interval; the discrete spectrum of the finite interval may pass into a continuous spectrum and the Fourier expansion in eigenfunctions may pass into a Fourier integral expansion for the infinite interval.

2. Bessel Functions. The situation is similar for the Bessel equation

$$(xu')' + \left(\lambda x - \frac{n^2}{x}\right)u = 0$$

in the interval $0 \leq x < \infty$, with the boundary condition that u remain finite for $x = 0$ and $x \to \infty$. All the Bessel functions $u = J_n(\sqrt{\lambda}\, x)$ with $\lambda \geq 0$ are eigenfunctions, so that we have a continuous spectrum consisting of all the non-negative values of λ.

Here, too, the expansion theorem for the representation of an arbitrary function $f(x)$ is replaced by an integral theorem, in which the domain of integration is the spectrum, i.e. the continuum of positive real numbers. This integral representation is

$$f(x) = \int_0^\infty t J_n(tx) g(t)\, dt, \qquad g(t) = \int_0^\infty \xi J_n(\xi t) f(\xi)\, d\xi.$$

Such a representation is possible if we assume that $f(x)$ is piecewise smooth for $x \geq 0$, that the integral

$$\int_0^\infty x \mid f(x) \mid dx$$

exists, and that $f(0) = 0$. The proof of this integral theorem will be given in Ch. VII, §2.

3. Eigenvalue Problem of the Membrane Equation for the Infinite Plane. The eigenvalue problem for the differential equation

$$\Delta u + \lambda u = 0$$

for the entire x, y-plane, with the boundary condition of boundedness at infinity, may be solved in two different ways.

The set of products $u = \sin\alpha(x - \xi)\sin\beta(y - \eta)$ of trigonometric functions, where ξ, η and α, β are arbitrary and the eigenvalues are the numbers $\lambda = \alpha^2 + \beta^2$, can be considered as eigenfunctions. An eigenvalue is then any non-negative number λ, and every such eigenvalue clearly determines a continuum of eigenfunctions. The corresponding integral representation is simply the Fourier integral theorem in the plane (cf. Ch. II, §6).

If we introduce polar coordinates r, φ, we obtain eigenfunctions of the form

$$u = J_n(\sqrt{\lambda}r)\sin n\varphi, \qquad u = J_n(\sqrt{\lambda}r)\cos n\varphi$$

where n is an arbitrary integer and λ any non-negative real number. In this case the spectrum is again the continuum of non-negative real numbers $\lambda \geq 0$; but, since n is an integer, only a countable number of eigenfunctions is associated with any eigenvalue $\lambda > 0$. Here the representation of an arbitrary function is obtained as a Fourier series expansion with respect to n and each coefficient is represented as an integral with respect to r, in accordance with the preceding subsection (see also Ch. VII, §2).

Incidentally these eigenfunctions are linear combinations of sine products corresponding to a given value of $\lambda = \alpha^2 + \beta^2$. In fact, we have

$$J_n(\sqrt{\lambda}r)e^{in\varphi} = \frac{(-i)^n}{2\pi}\int_0^{2\pi} e^{int}e^{ix\sqrt{\lambda}\cos t + iy\sqrt{\lambda}\sin t}\,dt,$$

(cf. Ch. VII, §2.)

4. The Schrödinger Eigenvalue Problem.[1] In his work on quantum theory, Schrödinger[2] was led to a type of eigenvalue problem with a

[1] Cf. also Ch. VI, §5.

[2] E. Schrödinger, Abhandlungen zur Wellenmechanik, Johann Ambrosius Barth, Leipzig, 1927.

spectrum of an entirely different structure. This spectrum consists of a continuous and a discrete part; the discrete part does not extend to infinity but has a finite point of accumulation. In the simplest Schrödinger problem we consider the eigenvalue equation

$$\Delta u + \frac{c}{r}u + \lambda u = 0, \tag{72}$$

where c is a given positive constant, and r, θ and φ are polar coordinates; we require that the eigenfunctions u be continuous at zero and remain finite as $r \to \infty$. If we multiply the differential equation by a spherical harmonic $Y_n(\theta, \varphi)$ and integrate over the unit sphere, we find in the usual way that the function

$$v(r) = \iint u(r, \theta, \varphi) Y_n(\theta, \varphi) \sin\theta \, d\theta \, d\varphi$$

satisfies the differential equation

$$v_{rr} + \frac{2}{r}v_r + \left(\lambda + \frac{c}{r} - \frac{n(n+1)}{r^2}\right)v = 0. \tag{73}$$

From the eigenfunctions of this equation, under the same boundary conditions as above for $r = 0$ and $r \to \infty$, we obtain eigenfunctions for (72) by forming the products $u = vY_n$.

Replacing λ by the new parameter

$$l = \frac{c}{2\sqrt{-\lambda}}$$

and r by the variable

$$z = 2\sqrt{-\lambda}\, r,$$

we are led to the differential equation

$$v_{zz} + \frac{2}{z}v_z + \left(-\frac{1}{4} + \frac{l}{z} - \frac{n(n+1)}{z^2}\right)v = 0,$$

which we encountered in a somewhat different form in §10, formula (51b). From the theory given there, we conclude: For real l, i.e. for negative λ, the condition that the solution be continuous at zero and remain finite as $r \to \infty$ can be satisfied only for integral values $l > n$; the solutions are given by the derivatives of the Laguerre polynomials in the form

$$v = z^n e^{-z/2} L_{l+n}^{(2n+1)}(z).$$

For the original equation we obtain the values

$$\lambda = -\frac{c^2}{4l^2}$$

(and only these) as the negative eigenvalues associated with the eigenfunctions

$$u = r^n e^{-cr/2l} L_{l+n}^{(2n+1)}\left(\frac{c}{l}r\right) Y_n(\theta, \varphi).$$

In this expression, for a given integer l, n can assume all integral values from 0 to $l - 1$, and Y_n denotes one of the $2n + 1$ linearly independent spherical harmonics. The discrete spectrum found in this way consists of infinitely many numbers with the limit point zero.

Furthermore, we state that Schrödinger's equation (72) *has all positive real numbers* λ *as eigenvalues*, i.e. *it has a continuous spectrum consisting of the non-negative real numbers.*

To prove this assertion, we introduce the function $w = rv$ in place of v in equation (73). This leads to the equation

$$w'' + \left(\lambda + \frac{c}{r} - \frac{n(n+1)}{r^2}\right)w = 0,$$

which is of the type treated in §11, 1. Accordingly, its solutions w remain bounded for all positive λ, and the solutions $v = w/r$ of equation (73) approach zero as r increases without bound. To demonstrate that every positive value of λ is an eigenvalue, we need only show that a solution v which is regular at the origin exists for all r. This fact is proved in the general theory of linear differential equations. However, a solution with the required properties may also be obtained directly, by a method which we have used before, in the form of an everywhere-convergent power series. This is best accomplished if we introduce $z = r^{-n}e^{ir\sqrt{\lambda}}v$ in the differential equation; thus we obtain an equation in z for which the assumption that the solution has a power series expansion leads to a two-term recursion formula.

§13. *Perturbation Theory*

If the eigenvalues λ_n and the associated orthonormal eigenfunctions u_n of the self-adjoint linear differential equation

$$L(u_n) + \lambda_n u_n = 0 \tag{74}$$

are known for a prescribed domain[1] and prescribed boundary conditions, then the eigenvalues and eigenfunctions of an eigenvalue problem corresponding to a "neighboring" or "perturbed" differential equation

$$L(\bar{u}_n) - \epsilon r\bar{u}_n + \bar{\lambda}_n \bar{u}_n = 0 \tag{75}$$

can be calculated by a method of approximations which is important in applications—the so-called *Calculus of Perturbations* or Perturbation Theory. It is understood that the boundary conditions and the domain remain unchanged. Let r denote a given function which is continuous in the fundamental domain and ϵ a parameter; $\bar{u}_n$ and $\bar{\lambda}_n$ are the eigenfunctions and eigenvalues of the new problem. Without proof we assume that the new eigenvalues as well as the new eigenfunctions may be expanded in powers of the perturbation parameter ϵ.

1. Simple Eigenvalues. Suppose, first, that the original unperturbed problem has only simple eigenvalues. Then we write

$$\bar{u}_n = u_n + \epsilon v_n + \epsilon^2 w_n + \cdots , \tag{76}$$

$$\bar{\lambda}_n = \lambda_n + \epsilon\mu_n + \epsilon^2 \nu_n + \cdots . \tag{77}$$

Insertion in (75) leads at once to equation (74) and to the additional equations

$$L(v_n) + \lambda_n v_n = ru_n - \mu_n u_n , \tag{78}$$

$$L(w_n) + \lambda_n w_n = rv_n - \mu_n v_n - \nu_n u_n . \tag{79}$$

From these, in turn, we can calculate the perturbations of various orders, i.e. the quantities μ_n, ν_n, $\cdots$ and v_n, w_n, $\cdots$, respectively.

In order to do this, we introduce the expansion coefficients

$$a_{nj} = \int v_n u_j \, dg$$

of the function v_n with respect to the eigenfunctions u_j. We multiply equation (78) by u_l, integrate over the fundamental domain, and transform the first term on the left by Green's formula using the boundary conditions (such as vanishing boundary values); we obtain

$$a_{nl}(\lambda_n - \lambda_l) = d_{nl} - \mu_n \delta_{nl} ,$$

[1] Here the number of dimensions of the domain is arbitrary. Integration is always to be carried out over the entire domain; the volume element is denoted by dg.

where $\delta_{nl} = 0$ for $n \neq l$ and $\delta_{nn} = 1$, and

$$d_{nl} = \int r u_n u_l \, dg.$$

By taking $l = n$ we obtain

$$\mu_n = d_{nn} \tag{80}$$

and, taking $l \neq n$,

$$a_{nl} = \frac{d_{nl}}{\lambda_n - \lambda_l}.$$

The quantity a_{nn} is found from the normalization condition $\int \bar{u}_n^2 \, dg = 1$, which leads to the relation $\int u_n v_n \, dg = 0$ and hence to $a_{nn} = 0$. Thus we find that if v_n can be expanded in terms of the u_j,

$$v_n = \sum_{j=1}^{\infty}{}' \frac{d_{nj}}{\lambda_n - \lambda_j} u_j \qquad (d_{nj} = \int r u_n u_j \, dg) \tag{81}$$

holds, where the sign $\sum'$ denotes summation over the indicated values of j omitting the term for which $j = n$.

Having determined the first approximation, we find the second

$$w_n = \sum_{j=1}^{\infty} b_{nj} u_j$$

in a similar way using equation (79), from which it follows, as above, that

$$b_{nl}(\lambda_n - \lambda_l) = \sum_{j=1}^{\infty} a_{nj} d_{jl} - \mu_n a_{nl} - \nu_n \delta_{nl}. \tag{82}$$

Setting $n = l$ we obtain the second perturbation term for the eigenvalue, namely

$$\nu_n = \sum_{j=1}^{\infty} a_{nj} d_{jn},$$

while for $l \neq n$ we find

$$b_{nl} = \frac{1}{\lambda_n - \lambda_l} \left\{ \sum_{j=1}^{\infty} a_{nj} d_{jl} - \mu_n a_{nl} \right\}. \tag{83}$$

To determine b_{nn} we must again employ the normalization condition $\int \bar{u}_n^2\, dg = 1$ and set the coefficient of ϵ^2 equal to zero. It follows easily that

$$b_{nn} = -\tfrac{1}{2} \sum_{j=1}^{\infty} a_{nj}^2 , \tag{84}$$

which completely determines the second approximation.

Further approximations may be determined successively in precisely the same manner.

2. Multiple Eigenvalues. A supplementary discussion is necessary in the case of multiple eigenvalues, the so-called "*degenerate*" case. It will be sufficient to assume that the first eigenvalue of (74) is α-fold, i.e. that $\lambda_1 = \cdots = \lambda_\alpha = \lambda$, while for $n > \alpha$ the remaining eigenvalues λ_n are simple. In the case of a multiple eigenvalue the eigenfunctions are determined only up to an orthogonal transformation. Thus, if the equation is perturbed, we cannot expect the individual eigenfunctions to vary continuously unless we have selected the system of unperturbed eigenfunctions for the multiple eigenvalue in a suitable way (see also Ch. III, §8, 4). Accordingly, we suppose the α eigenfunctions associated with the eigenvalue λ transformed into another system of such eigenfunctions by means of an orthogonal transformation

$$u_n^* = \sum_{j=1}^{\alpha} \gamma_{nj} u_j \qquad (n = 1, 2, \cdots, \alpha)$$

which will be determined later. We now assume

$$\bar{u}_n = u_n^* + \epsilon v_n + \epsilon^2 w_n + \cdots ,$$

and attempt to determine both the γ_{nj} and the functions v_n, w_n, $\cdots$. For $n > \alpha$ we have $u_n^* = u_n$, and the observations of subsection 1 remain valid. Thus we can limit ourselves to the cases $n = 1, 2, \cdots, \alpha$. Our assumption and equation (75) lead to the following equations for v_n and w_n, respectively:

$$L(v_n) + \lambda_n v_n = \sum_{j=1}^{\alpha} r\gamma_{nj} u_j - \mu_n \sum_{j=1}^{\alpha} \gamma_{nj} u_j , \tag{85}$$

$$L(w_n) + \lambda_n w_n = r v_n - \mu_n v_n - \nu_n \sum_{j=1}^{\alpha} \gamma_{nj} u_j . \tag{86}$$

If we multiply (85) by u_l and proceed as in subsection 1, using the same notation, we obtain

$$a_{nl}(\lambda_n - \lambda_l) = \sum_{j=1}^{\alpha} (d_{jl} - \mu_n \delta_{jl})\gamma_{nj}, \tag{87}$$

and hence in particular for $l = 1, 2, \cdots, \alpha$ we have

$$0 = \sum_{j=1}^{\alpha} (d_{jl} - \mu_n \delta_{jl})\gamma_{nj} \qquad (l, n = 1, 2, \cdots, \alpha).$$

From these α^2 equations, the quantities $\mu_1, \mu_2, \cdots, \mu_\alpha$ may be determined uniquely as roots of the characteristic equation $|d_{jl} - \mu_n \delta_{jl}| = 0$ (see Ch. I, §2). For simplicity, we may assume that these roots are all different, i.e. that the form $\sum_{j,l=1}^{\alpha} d_{jl} x_j x_l$ possesses only distinct eigenvalues. Then the orthogonal matrix (γ_{nj}) is also determined uniquely. This means that our choice of unperturbed eigenfunctions $u_n^* = \sum_{j=1}^{\alpha} \gamma_{nj} u_j$ is now fixed. Let us designate these u_n^* by u_n. The matrix (d_{nl}) in the new notation is a diagonal matrix with the elements

$$d_{nn} = \mu_n ;$$

the remaining elements are zero. From equations (87), for $l > \alpha$, we have immediately

$$a_{nl} = \frac{d_{nl}}{\lambda_n - \lambda_l}. \tag{88}$$

As in subsection 1, the normalization condition implies $a_{nn} = 0$; to determine the quantities a_{nl} $(l, n = 1, 2, \cdots, \alpha;\ n \neq l)$ we must use equations (86) for the second approximation. The latter, for $l, n = 1, 2, \cdots, \alpha$, lead to

$$0 = \sum_{j=1}^{\infty} a_{nj} d_{jl} - \mu_n a_{nl} - \nu_n \delta_{nl}$$

or, in view of the fact that (d_{jl}) $(j, l = 1, 2, \cdots, \alpha)$ is a diagonal matrix with the diagonal elements μ_n, to

$$0 = \sum_{j=\alpha+1}^{\infty} a_{nj} d_{jl} + a_{nl}\mu_l - \mu_n a_{nl} - \nu_n \delta_{nl}.$$

Thus when $l = n$ we have

$$\nu_n = \sum_{j=\alpha+1}^{\infty} a_{nj} d_{jn}, \tag{89}$$

a relation in which the coefficients a_{nj} are already determined by (88). For $n \neq l$ we obtain

$$a_{nl} = \frac{1}{\mu_n - \mu_l} \sum_{j=\alpha+1}^{\infty} a_{nj} d_{jl}.$$

We summarize the result: Given the α-fold degenerate multiple eigenvalue $\lambda_1 = \lambda$, select a system of orthonormal eigenfunctions $u_1, u_2, \cdots, u_\alpha$ for which the matrix $d_{nl} = \int r u_n u_l \, dg$ is a diagonal matrix with elements d_{nn}. Then the first order perturbation of the eigenvalue is given by

$$\mu_n = d_{nn}$$

and the first order perturbation of the eigenfunctions by

$$v_n = \sum_{j=1}^{\infty} a_{nj} u_j$$

with

$$a_{nn} = 0$$

and

$$a_{nl} = \frac{d_{nl}}{\lambda_n - \lambda_l},$$

if at least one of the indices l or n is greater than α, and by

$$a_{nl} = \frac{1}{d_{nn} - d_{ll}} \sum_{j=\alpha+1}^{\infty} \frac{d_{nj} d_{jl}}{\lambda_n - \lambda_j},$$

if neither index is greater than α and $l \neq n$.

The perturbation terms of second and higher orders are obtained in a corresponding manner. The second order perturbation of the n-th eigenvalues, in particular, is given by

$$\nu_n = \sum_{j=\alpha+1}^{\infty} \frac{d_{nj}^2}{\lambda_n - \lambda_j}$$

as we see from (89).

3. An Example.[1] Consider the problem of a freely vibrating string, with ends fixed at $x = 0$ and $x = \pi$, having the coefficient of elasticity

[1] Cf. Rayleigh, The Theory of Sound, Vol. I, pp. 115–118.

$p = 1$ and a mass density $\rho(x)$ which differs slightly from a constant value ρ_0 in the interval $0 \leq x \leq \pi$ and thus has the form $\rho(x) = \rho_0 + \epsilon\sigma(x)$, where $\sigma(x)$ is a given function and ϵ denotes a "perturbation parameter". By §3 the associated eigenvalue problem is of the form

$$\bar{u}_n'' + \bar{\lambda}_n(\rho_0 + \epsilon\sigma(x))\bar{u}_n = 0. \tag{90}$$

For $\epsilon = 0$ we get the undisturbed problem $u_n'' + \lambda_n\rho_0 u_n = 0$ which has the solution $\lambda_n = n^2/\rho_0$, $u_n = \sqrt{2/\pi\rho_0}\sin nx$.

Since all the eigenvalues are simple, we obtain the first approximation for the disturbed problem (90) by substituting the expressions[1]

$$\lambda_n = \frac{n^2}{\rho_0}, \qquad u_n = \sqrt{\frac{2}{\pi\rho_0}}\sin nx$$

and

$$r(x) = -\lambda_n\sigma(x) = -\frac{n^2}{\rho_0}\sigma(x)$$

in formulas (80) and (81) of subsection 1. Thus for the first order perturbation μ_n of the eigenvalues we obtain

$$\mu_n = -\frac{n^2}{\rho_0^2}\frac{2}{\pi}\int_0^\pi \sigma(x)\sin^2 nx\,dx$$

and for the eigenfunctions v_n we have

$$v_n = \sum_{j=1}^{\infty} a_{nj}u_j \tag{91}$$

where

$$a_{nj} = \frac{2}{\pi}\frac{n^2}{j^2 - n^2}\frac{1}{\rho_0}\int_0^\pi \sigma(x)\sin nx\sin jx\,dx \qquad (j \neq n); \tag{92}$$

$$a_{nn} = 0.$$

[1] In subsection 1 we assumed that the function $r(x)$ in the perturbation term $\epsilon r(x)$ did not depend on ϵ; however, the function $\bar{\lambda}_n\sigma(x)$ in the corresponding perturbation term of (90) depends on ϵ. Since we shall be concerned only with first order perturbations, we may set $r(x) = -\lambda_n\sigma(x)$, where λ_n no longer depends on ϵ.

In order to apply these results, we calculate the displacement δx of the first nodal point, which corresponds to $n = 2$ and lies at the midpoint in the case of a homogeneous string.

Since we have assumed that $\bar{u}_n$ can be expanded in powers of ϵ, we can write δx in the form $\delta x = \epsilon\tau + \epsilon^2(\cdots)$. We find the following equation for τ:

$$\begin{aligned} 0 &= \bar{u}_2\left(\frac{\pi}{2} + \epsilon\tau + \cdots\right) \\ &= u_2\left(\frac{\pi}{2} + \epsilon\tau + \cdots\right) + \epsilon v_2\left(\frac{\pi}{2} + \epsilon\tau + \cdots\right) + \epsilon^2(\cdots) \\ &= u_2\left(\frac{\pi}{2}\right) + \epsilon\left[\tau u_2'\left(\frac{\pi}{2}\right) + v_2\left(\frac{\pi}{2}\right)\right] + \epsilon^2(\cdots). \end{aligned}$$

If in this equation we set the coefficients of ϵ equal to zero and consider (91) we obtain

$$\tau = -\frac{v_2\left(\frac{\pi}{2}\right)}{u_2'\left(\frac{\pi}{2}\right)} = a_{21} - a_{23} + a_{25} - + \cdots,$$

since $u_2(x) = \text{const. } \sin 2x$. If, for example, the nonhomogeneity of the string is due to the presence of a small mass $\rho_0\kappa$ at the point $x = \pi/4$, we obtain from (92) by means of a simple limiting process the following expression for τ:

$$\begin{aligned} \tau &= \frac{4\kappa}{\pi}\left(\frac{\sin\frac{\pi}{4}}{1^2 - 4} - \frac{\sin\frac{3\pi}{4}}{3^2 - 4} + \frac{\sin\frac{5\pi}{4}}{5^2 - 4} - \cdots\right) \\ &= \frac{4\kappa}{\pi\sqrt{2}}\left(\frac{1}{1^2 - 4} - \frac{1}{3^2 - 4} - \frac{1}{5^2 - 4} + \cdots\right) \\ &= -\frac{2\kappa}{\pi\sqrt{2}}\left(1 + \frac{1}{3} - \frac{1}{5} - \frac{1}{7} + \frac{1}{9} + \frac{1}{11} - \cdots\right). \end{aligned}$$

The value of the series in parentheses is

$$\int_0^1 \frac{1 + x^2}{1 + x^4}\,dx = \frac{\pi}{4}\sqrt{2}$$

and $\tau = -\kappa/2$.

§14. *Green's Function (Influence Function) and Reduction of Differential Equations to Integral Equations*

In this section we shall consider and supplement the theory from an essentially different point of view. Instead of beginning with a vibration or eigenvalue problem, we shall start with a boundary value problem and develop the method of Green's function (or influence function) to represent its solutions. In this way, the eigenvalue differential equations will be reduced to symmetric integral equations; the existence of the eigenfunctions and validity of the completeness and expansion theorems will thereby be automatically proved.

1. Green's Function and Boundary Value Problem for Ordinary Differential Equations. First we consider a linear homogeneous self-adjoint differential expression of second order

$$L[u] = pu'' + p'u' - qu$$

for the function $u(x)$ in the fundamental domain G: $x_0 \leq x \leq x_1$, where p, p' and q are continuous functions of x and $p > 0$. The associated nonhomogeneous differential equation is of the form

$$L[u] = -\varphi(x), \tag{93}$$

where $\varphi(x)$ denotes a piecewise continuous function defined in G. We are concerned with the boundary value problem: find a solution $u = f(x)$ of equation (93) which satisfies given homogeneous boundary conditions at the boundary of G, e.g., the boundary condition $u = 0$.[1] It is natural to start with the following heuristic considerations: We again consider equation (93) as the condition for equilibrium of a string under the influence of a time-independent force distributed continuously with density $\varphi(x)$ over the string. We visualize a limiting process from the continuously distributed force to a "point force," i.e. to a force acting at a single point $x = \xi$ with a given intensity. The string is subjected to boundary conditions. Let $K(x, \xi)$ denote the deflection of the string at a point x as a result of the action of a point force of unit intensity applied at the point ξ; then the effect at x of the continuously distributed force φ can be considered as the superposition of the effects of continuously distributed point forces

[1] It will be recalled that the boundary value problem of the homogeneous differential equation with nonhomogeneous boundary conditions can always be reduced to this problem. (Cf. §1, 2).

whose linear density at each point ξ is equal to $\varphi(\xi)$. Thus we can expect the desired solution to be of the form

$$u(x) = \int_{x_0}^{x_1} \mathsf{K}(x, \xi)\varphi(\xi)\, d\xi. \tag{94}$$

The function $\mathsf{K}(x, \xi)$, called the *influence function* or *Green's function* for the differential expression $L[u]$, satisfies the prescribed conditions at $x = x_0$ and $x = x_1$. It follows that the function $u(x)$, which is represented in equation (94) by an integral in terms of the kernel $\mathsf{K}(x, \xi)$ and the density $\varphi(x)$, also satisfies these boundary conditions.

The influence function $\mathsf{K}(x, \xi)$ satisfies the differential equation

$$L[\mathsf{K}] = 0$$

everywhere except at the point $x = \xi$, since K corresponds to a force zero when $x \neq \xi$. At the point $x = \xi$ the function $\mathsf{K}(x, \xi)$ must have a singularity, which can be found heuristically in the following way: We consider the point force as the limiting case of a force $\varphi_\epsilon(x)$ which vanishes in G for $| x - \xi | > \epsilon$ and the total intensity of which is given by the equation

$$\int_{\xi-\epsilon}^{\xi+\epsilon} \varphi_\epsilon(x)dx = 1.$$

Denote the associated deflection of the string by $\mathsf{K}_\epsilon(x, \xi)$; thus, $L[\mathsf{K}_\epsilon] = (p\mathsf{K}_\epsilon')' - q\mathsf{K}_\epsilon = -\varphi_\epsilon(x)$. Integrating this equation between the limits $\xi - \delta$ and $\xi + \delta$, where $\delta \geq \epsilon$ may be chosen arbitrarily provided the interval of integration remains in the fundamental domain G, we obtain

$$\int_{\xi-\delta}^{\xi+\delta} \left(\frac{d}{dx}\left(p \frac{d\mathsf{K}_\epsilon}{dx} \right) - q\mathsf{K}_\epsilon \right) dx = -1.$$

If we first take the limit as $\epsilon \to 0$, assuming that K_ϵ converges to a continuous function $\mathsf{K}(x, \xi)$ which is continuously differentiable except at $x = \xi$, and then let δ approach zero, we obtain the relation

$$\lim_{\delta \to 0} \frac{d\mathsf{K}(x, \xi)}{dx} \Bigg|_{x=\xi-\delta}^{x=\xi+\delta} = -\frac{1}{p(\xi)},$$

which characterizes the singularity of Green's function.

Reversing this heuristic discussion, we shall turn it into a rigorous mathematical theory. We begin by *defining* a function $\mathsf{K}(x, \xi)$ as *Green's function* of the differential expression $L[u]$, under

given homogeneous boundary conditions; we stipulate the following properties:

(1) $\mathsf{K}(x, \xi)$ is, for fixed ξ, a continuous function of x and satisfies the prescribed boundary conditions.

(2) Except at the point $x = \xi$, the first and second order derivatives of K with respect to x are continuous in G. At the point $x = \xi$ the first derivative has a jump discontinuity given by

$$\frac{d\mathsf{K}(x, \xi)}{dx}\Bigg|_{x=\xi-0}^{x=\xi+0} = -\frac{1}{p(\xi)}. \tag{95}$$

(3) K considered as a function of x satisfies the differential equation $L[\mathsf{K}] = 0$ throughout G, except at the point $x = \xi$.

A continuous function which satisfies conditions (2) and (3), but not necessarily the boundary conditions, is called a "*fundamental solution*" of the differential equation $L[u] = 0$.

We shall prove the following relations, which seem plausible in view of the above discussion: *If $\varphi(x)$ is a continuous or piecewise continuous function of x, then the function*

$$u(x) = \int_{x_0}^{x_1} \mathsf{K}(x, \xi)\varphi(\xi)\, d\xi \tag{96}$$

is a solution of the differential equation

$$L[u] = -\varphi(x) \tag{97}$$

and satisfies the boundary conditions. Conversely, if the function $u(x)$ satisfies equation (97) *and the boundary conditions, it can be represented by* (96). To prove the first statement only the elementary rules for the differentiation of an integral with respect to a parameter are needed. Thus, because of (95), we obtain consecutively the following equations:

$$u'(x) = \int_{x_0}^{x_1} \mathsf{K}'(x, \xi)\varphi(\xi)\, d\xi,$$

$$\begin{aligned} u''(x) &= \int_{x_0}^{x} \mathsf{K}''(x, \xi)\varphi(\xi)\, d\xi + \int_{x}^{x_1} \mathsf{K}''(x, \xi)\varphi(\xi)\, d\xi \\ &\quad + \mathsf{K}'(x, x - 0)\varphi(x) - \mathsf{K}'(x, x + 0)\varphi(x) \\ &= \int_{x_0}^{x_1} \mathsf{K}''(x, \xi)\varphi(\xi)\, d\xi + (\mathsf{K}'(x + 0, x) - \mathsf{K}'(x - 0, x))\varphi(x) \\ &= \int_{x_0}^{x_1} \mathsf{K}''(x, \xi)\varphi(\xi)\, d\xi - \frac{\varphi(x)}{p(x)}; \end{aligned}$$

therefore,

$$pu'' + p'u' - qu = \int_{x_0}^{x_1} (p\mathsf{K}'' + p'\mathsf{K}' - q\mathsf{K})\varphi(\xi)\, d\xi - \varphi(x).$$

This completes the proof, since $L[\mathsf{K}] = 0$.

To prove the converse we again use Green's formula (§1, (2b)); setting $v = \mathsf{K}$ we apply the formula to the two domains of integration $x_0 \leq x \leq \xi$ and $\xi \leq x \leq x_1$. Formula (96), with x and ξ interchanged, then follows directly from the discontinuity relation and the boundary conditions.

If u and K are not subjected to the same homogeneous boundary conditions (e.g. $\mathsf{K} = 0$ at the end points $x = x_0$ and $x = x_1$), we obtain in the same way the more general expression for u

$$u(x) = \int_{x_0}^{x_1} \mathsf{K}(x, \xi)\varphi(\xi)\, d\xi + p\mathsf{K}'u \Big|_{x_0}^{x_1};$$

for $\varphi = 0$ this expression furnishes the representation of the solution of the boundary value problem for the homogeneous differential equation $L[u] = 0$ in terms of the boundary values.

Green's function of a self-adjoint differential expression is a symmetric function of the parameter ξ and the argument x, that is,

$$\mathsf{K}(x, \xi) = \mathsf{K}(\xi, x).$$

This follows almost immediately from Green's formula (§1, (2b)), if we substitute $v = \mathsf{K}(x, \eta)$, $u = \mathsf{K}(x, \xi)$ and divide the domain of integration into the three intervals $x_0 \leq x \leq \xi$, $\xi \leq x \leq \eta$, $\eta \leq x \leq x_1$, treating each interval separately. The proof is completed by taking into account both the discontinuity relation (95) at the points $x = \xi$ and $x = \eta$ and the boundary conditions. *The symmetry of Green's function expresses a reciprocity frequently occurring in physics: If the force* 1, *applied at the point ξ, produces the result $\mathsf{K}(x, \xi)$ at the point x, then the force* 1 *acting at x produces the same result at ξ.*

2. Construction of Green's Function; Green's Function in the Generalized Sense. We construct Green's function for $L[u]$ under the prescribed boundary conditions: Consider any solution $u_0(x)$ of the differential equation $L[u] = 0$ which at $x = x_0$ satisfies the given boundary condition, e.g. vanishes. Then $c_0u_0(x)$ is the most general such solution. Similarly, let $c_1u_1(x)$ be the family of solutions of $L[u] = 0$ which satisfy the boundary condition at $x = x_1$. There

are two possible cases: Either the two families of solutions are distinct—the "general" case—or they are identical. In the first case the functions u_0, u_1 are linearly independent, in other words, $u_0u_0' - u_1'u_1 \not\equiv 0$.[1] In this case a curve of the first family can never touch a curve of the second family in the fundamental domain (for, if it did, the equation would be contradicted at the point of contact). We can choose the two constants c_0, c_1 such that the point of intersection has a given abscissa $x = \xi$ in the interval G and such that the discontinuity of the derivative at this point has precisely the value $-1/p(\xi)$. In this fashion we obtain Green's function $K(x, \xi)$ explicitly by means of the formulas

$$x \leq \xi: \qquad u = -\frac{1}{c}\, u_1(\xi)u_0(x)$$

$$x \geq \xi: \qquad u = -\frac{1}{c}\, u_0(\xi)u_1(x)$$

$$c = p(\xi)[u_0(\xi)u_1'(\xi) - u_0'(\xi)u_1(\xi)] = \text{const.}$$

In the second case, u_0 and u_1 differ only by a constant factor. Every solution belonging to one family also belongs to the other. Thus in this case the function $u_0(x)$ satisfies not only the condition for $x = 0$ but also the condition for $x = 1$, and the equation $L[u] = 0$ has a nontrivial solution $u_0(x)$ that satisfies the boundary conditions. This can also be expressed by stating: $\lambda = 0$ is an eigenvalue of $L[u] + \lambda u = 0$. Hence the above construction fails, and no Green's function exists.

The existence of Green's function is equivalent to the existence of a unique solution of the homogeneous boundary value problem for the differential equation $L[u] = -\varphi(x)$ (cf. subsection 1). Therefore the following *alternative* exists: *Under given homogeneous boundary conditions, either the equation* $L[u] = -\varphi(x)$ *has a uniquely determined solution* $u(x)$ *for every given* φ, *or the homogeneous equation* $L[u] = 0$ *has a nontrivial solution.*

Moreover, it will turn out that in the second case *the nonhomogeneous equation* $L[u] = -\varphi(x)$ *has a solution if, and only if, the*

[1] This is immediately clear. For, if c is a constant, then $\Delta = u_0u_1' - u_0'u_1 = c/p$ holds, as can be easily verified by deriving the equation $p\Delta' + \Delta p' = 0$ from the given differential equation.

solutions $u_0(x)$ of the homogeneous equation $L[u_0] = 0$ are orthogonal to the function φ as expressed in the orthogonality condition

$$\int_{x_0}^{x_1} \varphi(x)u_0(x)\,dx = 0.$$

The orthogonality condition is seen to be necessary if we multiply the differential equation $L[u] + \varphi(x) = 0$ by the function $u_0(x)$, integrate over the domain G, and apply Green's formula, taking into account the boundary conditions. That the condition is also sufficient can be shown if we introduce in place of the Green's function a "*Green's function in the generalized sense.*" We are led to this function again by a simple heuristic argument based on intuition. We recall: If λ is an eigenvalue with an associated normalized eigenfunction u, then under the influence of an external force of the form $-\psi(x)e^{i\sqrt{\lambda}t}$ the string becomes unstable (resonance) unless the relation $\int_{x_0}^{x_1} \psi(x)u(x)\,dx = 0$ is satisfied. In the present case, where $\lambda = 0$, this means instability under the influence of an external time-independent force. In particular, the string is unstable when a point force is applied at an arbitrary point. If the system is not to depart arbitrarily far from its rest state when a point force is applied, then the string must be balanced by a fixed time-independent opposing force. This opposing force may be chosen arbitrarily, except that it may not be orthogonal to the eigenfunction $u_0(x)$, since then it would not prevent the excitation of the eigenfrequency zero. It is convenient to assume that this balancing force is in the symmetric form $\psi(x) = -u_0(x)u_0(\xi)$. Then the influence function $\mathsf{K}(x, \xi)$ of a point force acting at the point $x = \xi$ satisfies not only the boundary conditions but also, except at the point $x = \xi$, the differential equation

$$L[\mathsf{K}] = u_0(x)u_0(\xi),$$

and at $x = \xi$ it must satisfy the discontinuity condition (95). These conditions determine a function K only up to an arbitrary additive function $c(\xi)u_0(x)$. To single out a particular function K, we require

$$\int_{x_0}^{x_1} \mathsf{K}(x, \xi)u_0(x)\,dx = 0$$

and call the function K so defined Green's function in the generalized sense for the differential expression $L[u]$. Using the assumption that

$L[u]$ is a self-adjoint differential expression, we find, as before, that Green's function in the generalized sense has the *symmetry property*

$$K(x, \xi) = K(\xi, x).$$

These considerations may be illustrated by a simple example, the string with both ends free (see also §15, 1). Here $u_0 = \text{const.}$ is an eigenfunction for $\lambda = 0$; for the opposing force we take a force which is constant along the entire length of the string.

Green's function in the generalized sense can be constructed in the same way as the ordinary Green's function. We have only to prove the fact: if $L[u] = 0$ has a nontrivial solution u_0 which satisfies the boundary conditions, then $L[v] = u_0(\xi)u_0(x)$ can have no such solution. In fact, if we multiply the latter equation by $u_0(x)$ and integrate over the fundamental domain, taking account of the boundary conditions, we obtain

$$u_0(\xi) \int_{x_0}^{x_1} u_0^2(x)\, dx = \int_{x_0}^{x_1} v(x) L[u_0]\, dx = 0,$$

which contradicts the assumption that $\int_{x_0}^{x_1} u_0^2(x)\, dx \neq 0$.

Green's function in the generalized sense serves the same purpose as the ordinary Green's function. Note that the solution $w(x)$ of a differential equation $L[w] = -\varphi(x)$ is determined only up to an arbitrary additive function $cu_0(x)$; it can therefore be fixed by means of the condition $\int_{x_0}^{x_1} wu_0\, dx = 0$. We can thus state the theorem: *Let $w(x)$ be a function which is orthogonal to $u_0(x)$, satisfies the boundary conditions, and has continuous first and piecewise continuous second derivatives. If $w(x)$ and the piecewise continuous function $\varphi(x)$ are related by the equation*

$$L[w] = -\varphi(x),$$

then relation

$$w(x) = \int_{x_0}^{x_1} K(x, \xi)\varphi(\xi)\, d\xi. \tag{98}$$

also holds.

Conversely, the latter relation implies the former, if $\varphi(x)$ is orthogonal to $u_0(x)$. This converse contains the second part of the theorem stated above (p. 355).

The proof is similar to the corresponding proof for the ordinary Green's function, but every function $w(x)$ of the form (98) must be orthogonal to $u_0(x)$, since this is true of Green's function $\mathsf{K}(x, \xi)$.

As we have already seen, $\lambda = 0$ can be at most a simple eigenvalue in the case of the second order differential equations considered here. If, however, $\lambda = 0$ is a multiple eigenvalue (for differential equations of higher order), one can construct Green's function in the generalized sense with an opposing force of the form

$$\psi(x) = -u_0(x)u_0(\xi) - u_1(x)u_1(\xi) - \cdots$$

where $u_0, u_1, \cdots$ denote the orthonormal eigenfunctions belonging to the eigenvalue $\lambda = 0$.

3. Equivalence of Integral and Differential Equations. By means of Green's function the eigenvalue problems discussed earlier can be completely solved if we replace the differential equation by an integral equation. We consider a linear family of differential equations

$$L[u] + \lambda\rho u = \psi(x) \qquad (\rho(x) > 0), \tag{99}$$

depending on a parameter λ; here $\psi(x)$ is a piecewise continuous function, $\rho(x)$ a positive continuous function, and u, by assumption, satisfies the given boundary conditions, say $u = 0$. If Green's function for $L[u]$ exists under the given boundary conditions, the following equation is immediately obtained from formula (94) by setting $\varphi(x) = \lambda\rho u - \psi$:

$$u(x) = \lambda \int_{x_0}^{x_1} \mathsf{K}(x, \xi)\rho(\xi)u(\xi)\, d\xi + g(x), \tag{100}$$

where

$$g(x) = -\int_{x_0}^{x_1} \mathsf{K}(x, \xi)\psi(\xi)\, d\xi$$

is a given continuous function of x; equation (100) is equivalent to (99). Hence, finding the desired solution u of (99) under the prescribed boundary conditions is equivalent to solving the integral equation (100). To the homogeneous equation

$$L[u] + \lambda\rho u = 0 \tag{101}$$

corresponds the homogeneous integral equation

$$u(x) = \lambda \int_{x_0}^{x_1} \mathsf{K}(x, \xi)\rho(\xi)u(\xi)\, d\xi.$$

If we introduce

$$u(x)\sqrt{\rho(x)} = z(x)$$

as a new unknown function, multiply the integral equation by $\sqrt{\rho(x)}$, and set $K(x, \xi) = \mathsf{K}(x, \xi)\sqrt{\rho(x)\rho(\xi)}$, then the integral equation

$$z(x) = \lambda \int_{x_0}^{x_1} K(x, \xi) z(\xi)\, d\xi \tag{102}$$

corresponds to (101). The kernel $K(x, \xi)$ of equation (102) is symmetric, since $L[u]$ is self-adjoint.[1] We can therefore apply the theorems on symmetric kernels investigated in Chapter III and obtain results—partly contained in subsection 2—for the differential equation (99).

The following alternatives hold for the relation between the boundary value problem of the nonhomogeneous differential equation (99) and that of the homogeneous differential equation (101) under the given homogeneous boundary conditions: *Either, for fixed* λ, *every solution of the homogeneous differential equation* (101) *vanishes identically* (λ *is not an eigenvalue of* (101)); *then the nonhomogeneous equation* (99) *has one and only one solution for arbitrarily chosen* $\psi(x)$. *Or, for some value* $\lambda = \lambda_i$, *the homogeneous equation* (101) *has a nontrivial solution* u_i (λ_i *is an eigenvalue of* (101) *with the eigenfunction* u_i); *then a solution of the nonhomogeneous differential equation* (99) *exists for* $\lambda = \lambda_i$ *if and only if the relation*

$$\int_{x_0}^{x_1} \rho u_i \psi\, dx = 0$$

holds for all eigenfunctions u_i *associated with the eigenvalue* λ_i.

Moreover: *There exists a sequence of eigenvalues* $\lambda_1, \lambda_2, \cdots$ ($\lambda_n \to \infty$)—*with associated eigenfunctions* $u_1, u_2, \cdots$ *which form an infinite system of functions satisfying the orthogonality relations*

$$\int_{x_0}^{x_1} \rho u_i u_k\, dx = 0 \qquad (i \neq k), \qquad \int_{x_0}^{x_1} \rho u_i^2\, dx = 1.$$

If, with Green's function $\mathsf{K}(x, \xi)$ *as kernel, we can represent a function* $w(x)$ *by a piecewise continuous function* $\varphi(\xi)$ *in the form*

$$w(x) = \int_{x_0}^{x_1} \mathsf{K}(x, \xi)\varphi(\xi)\, d(\xi),$$

[1] This symmetry and its consequences show the significance of the assumption that $L[u]$ is self-adjoint.

then $w(x)$ can be expanded in terms of the eigenfunctions into an absolutely and uniformly convergent series

$$w(x) = \sum_{n=1}^{\infty} c_n u_n(x), \qquad c_n = \int_{x_0}^{x_1} w\rho u_n \, dx.$$

The set of functions which can be expanded in this way may be more simply characterized. The equation $L[w] = -\varphi(x)$ follows from (94) because of the fundamental property of Green's function. Conversely, if we consider any function $w(x)$ which satisfies the boundary conditions and has a continuous first and a piecewise continuous second derivative, we can construct a corresponding source-distribution $\varphi(x)$ by means of the equation $L[w] = -\varphi(x)$. Thus we obtain the result: *Every function $w(x)$ which satisfies the boundary conditions and has a continuous first and a piecewise continuous second derivative may be expanded in an absolutely and uniformly convergent series $w(x) = \sum_{n=1}^{\infty} c_n u_n(x)$.*

This theorem implies immediately *that the eigenfunctions form a complete orthogonal system.* For, any function continuous in G may be approximated in the mean to any desired degree of accuracy by a continuous function which satisfies the boundary conditions and has continuous first and second derivatives. Thus, in virtue of the expansion theorem just stated, it may also be approximated by a finite combination of the form $\sum_{n=1}^{m} c_n u_n(x)$.

The fact, noted earlier,[1] that all the eigenvalues are positive (in the terminology of the theory of integral equations, this means that the kernel $K(x, \xi)$ is definite) leads to a *stronger form of the expansion theorem.* Since, in addition, $K(x, \xi)$ is a continuous function of x and ξ, we can apply Mercer's theorem of Ch. III, §5, 4 and conclude that the series expansions of the kernel

$$K(x, \xi) = \sqrt{\rho(x)\rho(\xi)} \sum_{n=1}^{\infty} \frac{u_n(x)u_n(\xi)}{\lambda_n} \tag{103}$$

$$\text{or} \qquad \mathsf{K}(x, \xi) = \sum_{n=1}^{\infty} \frac{u_n(x)u_n(\xi)}{\lambda_n}$$

converge absolutely and uniformly. This formula, explicitly relating Green's function to eigenfunctions, is called simply the *bilinear relation*; for fixed ξ it represents a series expansion of a continuous func-

[1] Cf. p. 294.

tion with a piecewise continuous first derivative. If we form a linear combination

$$S = \alpha_1 K(x, \xi_1) + \alpha_2 K(x, \xi_2) + \cdots,$$

we obtain a continuous function the first derivative of which has the preassigned jumps $-\alpha_1/p(\xi_1)$, $-\alpha_2/p(\xi_2)$, $\cdots$ at the prescribed points ξ_1, ξ_2, $\cdots$; this function can be expanded in an absolutely and uniformly convergent eigenfunction series. Since from any function with piecewise continuous first and second derivatives we can subtract a particular function S in such a way that the difference satisfies the conditions of the above expansion theorem, we obtain the result: *A sufficient condition for the validity of the expansion theorem is that the first and second derivatives of the continuous function $w(x)$ should be piecewise continuous.*

In this section we have assumed that Green's function for $L[u]$ exists, i.e. that $\lambda = 0$ is not an eigenvalue of our differential equation $L[u] + \lambda \rho u = 0$ (subsection 2). If this assumption is not valid, we replace the ordinary Green's function by Green's function in the generalized sense; all considerations leading to the reduction of the eigenvalue problem of equation (101) to an integral equation problem hold. To insure the validity of the expansion theorem we must add the condition of orthogonality to the eigenfunction $u_0(x)$ associated with $\lambda = 0$. This condition, however, disappears completely from the final formulation of the expansion theorem if we include the eigenfunctions associated with the eigenvalue $\lambda = 0$. The occurrence of a zero eigenvalue is in no way pathological; this will be seen in connection with an approach to the eigenvalue problem by a method based on the calculus of variations (cf. Ch. VI, §1).

In conclusion, we shall expand the solution of the nonhomogeneous equation (99) with respect to the eigenfunctions. Corresponding to the earlier scheme of §3, 3, which can now be justified by the expansion theorem, or directly by the integral equation theorem (Chapter III, formula (56)) we obtain the solution

$$u(x) = \sum_{n=1}^{\infty} \gamma_n u_n(x) \quad \text{with} \quad \gamma_n = \frac{c_n}{\lambda - \lambda_n}, \qquad c_n = \int_{x_0}^{x_1} u_n(x)\psi(x)\, dx.$$

This clarifies the fact that, if $\lambda = \lambda_i$ is an eigenvalue, equation (99) cannot be solved unless the orthogonality relation $\int_{x_0}^{x_1} \psi u_i\, dx = 0$ is

satisfied. In physical terms: *When the external force is in resonance with an eigenvibration, a stationary state exists if and only if this force performs no work on the system as the system moves in the given eigenvibration.*

4. Ordinary Differential Equations of Higher Order. Ordinary differential equations of higher order are not essentially different. We restrict the discussion to a typical example, associated with the differential equation $u'''' - \lambda u = 0$ or $u'''' - \lambda \rho u = 0$, the homogeneous or nonhomogeneous rod (cf. §4). As before, we introduce the influence function or Green's function $\mathsf{K}(x, \xi)$ as the displacement of the rod, in the state of equilibrium, under the influence of a point force acting at the point $x = \xi$ and satisfying the prescribed boundary conditions. In the same manner as above we obtain the following typical conditions for this function:

(1) For every value of the parameter ξ the function $\mathsf{K}(x, \xi)$, together with its first and second derivatives, is continuous and satisfies the prescribed homogeneous boundary conditions.

(2) At any point different from $x = \xi$ the third and fourth derivatives with respect to x are also continuous. However, at $x = \xi$ the following discontinuity condition holds:

$$\lim_{\epsilon \to 0} [\mathsf{K}'''(\xi + \epsilon, \xi) - \mathsf{K}'''(\xi - \epsilon, \xi)] = -1.$$

(3) Except at the point $x = \xi$, the differential equation

$$\mathsf{K}''''(x, \xi) = 0$$

is satisfied everywhere.

The fundamental property of Green's function can be stated as follows: Let $u(x)$ be a continuous function which satisfies the boundary conditions and has continuous first, second, and third, and piecewise continuous fourth derivatives; let $\varphi(x)$ be a piecewise continuous function. If $u(x)$ and $\varphi(x)$ are connected by the relation

$$L[u] = u'''' = -\varphi(x),$$

then we have the representation

$$u(x) = \int_{x_0}^{x_1} \mathsf{K}(x, \xi)\varphi(\xi)\, d\xi$$

and conversely.

The eigenvalue problem of the more general differential equation

$$u'''' - \lambda \rho u = 0,$$

the associated expansion theorem, and the theory of the nonhomogeneous equation

$$u'''' - \lambda \rho u = -\psi(x)$$

may be treated like the corresponding questions in subsection 3, i.e. by reduction to an integral equation with the symmetric kernel $K(x, \xi) = \mathsf{K}(x, \xi)\sqrt{\rho(x)\rho(\xi)}$. The result obtained is: *There exists an infinite system of eigenvalues* $\lambda_1, \lambda_2, \cdots$ *and associated eigenfunctions* $u_1, u_2, \cdots$ *with the properties that the functions* $\sqrt{\rho}u_i$ *constitute a complete orthogonal system and that any function* $w(x)$ *which satisfies the boundary conditions and has continuous derivatives up to the third order and a piecewise continuous fourth derivative may be expanded in terms of these functions in an absolutely and uniformly convergent series. Moreover, Mercer's theorem*[1] *implies that the bilinear relation*

$$\mathsf{K}(x, \xi) = \sum_{n=1}^{\infty} \frac{u_n(x)u_n(\xi)}{\lambda_n}$$

holds and that the expansion theorem can be extended to functions whose third derivatives are only piecewise continuous.

The existence proof and construction of Green's function and of Green's function in the generalized sense offer no new difficulties. They will be illustrated by examples in §15.

5. Partial Differential Equations. In the case of partial differential equations with homogeneous boundary conditions, Green's function can again be introduced as the kernel of an equivalent integral equation. As an example, consider the second order partial differential equation

$$\Delta v = -\varphi(x, y)$$

in the x, y-plane for a domain G under a homogeneous boundary condition, e.g. $v = 0$. This equation characterizes the deflection of a stretched membrane in a state of equilibrium under the influence of a time-independent force of density $\varphi(x, y)$. As before, the solution can be obtained with the aid of a Green's function $\mathsf{K}(x, y; \xi, \eta)$, repre-

[1] Cf. Ch. III, §5, 4. As in the problem of the vibrating string (cf. p. 360) we can show that the kernel is positive definite.

senting the influence at the point (x, y) of a point force acting at the source point (ξ, η). K and its derivatives of first and second order must be continuous everywhere except at the point $x = \xi, y = \eta$, and the differential equation $\Delta\mathsf{K} = 0$ must be satisfied. Furthermore, K must satisfy the given homogeneous boundary conditions, and possess a characteristic singularity of a point force at the *source point* $x = \xi$, $y = \eta$. The nature of this singularity is determined if we surround the source point with a circle k of radius ϵ, assume an external force of density $\varphi_\epsilon(x, y)$ vanishing outside k for which $\iint_k \varphi_\epsilon(x, y)dx\,dy = 1$, and consider the Green's function $\mathsf{K}(x, y; \xi, \eta)$ as the limit (for $\epsilon \to 0$) of that solution $\mathsf{K}_\epsilon(x, y; \xi, \eta)$ of $\Delta\mathsf{K} = -\varphi_\epsilon$ which satisfies the given boundary conditions. If we integrate the equation $\Delta\mathsf{K} = -\varphi_\epsilon$ over the circle of radius $\delta \geq \epsilon$ and apply Green's formula (5a) of page 280, we obtain

$$\int_\kappa \frac{\partial}{\partial r}\mathsf{K}_\epsilon\, ds = -1.$$

Here $r = \sqrt{(x-\xi)^2 + (y-\eta)^2}$ denotes the distance of the point (x, y) from the point (ξ, η), and s is the arc length measured on the boundary κ. We shall therefore subject the Green's function which we wish to characterize to the condition

$$\int_\kappa \frac{\partial}{\partial r}\mathsf{K}(x, y; \xi, \eta)\, ds = -1.$$

This condition is satisfied if we require that, in a neighborhood of the source point, K be of the form

$$\mathsf{K}(x, y; \xi, \eta) = -\frac{1}{2\pi}\log r + \gamma(x, y; \xi, \eta),$$

where $\gamma(x, y; \xi, \eta)$, together with its derivatives of first and second order, is continuous in x and y. (Since $\log r$ is, for $r \neq 0$, a solution of the potential equation, γ itself is a regular potential function.)

Reversing the order of this heuristic discussion, we define Green's function K by means of the following conditions:

(1) Except at the source point (ξ, η) the function $\mathsf{K}(x, y; \xi, \eta)$ and its derivatives of first and second order with respect to x and y are continuous. The function K has the form

$$\mathsf{K}(x, y; \xi, \eta) = -\frac{1}{2\pi}\log r + \gamma(x, y; \xi, \eta),$$

where $\gamma(x, y; \xi, \eta)$ and its derivatives up to the second order are continuous.

(2) K satisfies the prescribed homogeneous boundary conditions.

(3) Everywhere except at the source point, the differential equation $\Delta K = 0$ is satisfied.

Green's function so defined satisfies the *symmetry condition*

$$K(x, y; \xi, \eta) = K(\xi, \eta; x, y).$$

The proof of this symmetry, which is another expression of the physical *reciprocity* noted above, is again obtained from Green's formula. We apply this formula for the functions $K(x, y; \xi, \eta)$ and $K(x, y; \xi', \eta')$ to a domain which is obtained from G by removing the circular disks k and k', each of radius ϵ, drawn about the points (ξ, η) and (ξ', η'), respectively. Taking the limit $\epsilon \to 0$ and keeping in mind the singularity property of Green's function, we obtain—since the integral over the boundary Γ of G vanishes because of the boundary conditions—the symmetry formula in the form $K(\xi', \eta'; \xi, \eta) = K(\xi, \eta; \xi', \eta')$.

As before, the fundamental property of Green's function is: *Let $u(x,y)$ be any function which satisfies the homogeneous boundary conditions—say $u = 0$—and which is continuous and has continuous first and piecewise continuous second derivatives in G. If*

$$L[u] = \Delta u = -\varphi(x, y),$$

the relation

$$u(x, y) = \iint_G K(x, y; \xi, \eta)\varphi(\xi, \eta)\, d\xi\, d\eta$$

holds. On the other hand, if $\varphi(x, y)$ is any function which, with its first derivatives, is continuous in G, then the function

$$u(x, y) = \iint_G K(x, y; \xi, \eta)\varphi(\xi, \eta)\, d\xi\, d\eta,$$

which is continuous in G, has continuous first and second derivatives and satisfies both the boundary condition and the differential equation

$$\Delta u = -\varphi(x, y).$$

Note that the differentiability assumption for $\varphi(x,y)$ is more stringent in the second part of the theorem than in the first. (For ordinary differential equations, there was no such difference.)

The first part of the theorem follows again almost immediately from Green's formula (5a). For $v = \mathsf{K}(x, y; \xi, \eta)$ we apply this formula to the domain $G - k$, which is obtained from G by removing a small circular disk k of radius ϵ about the point (x, y); again we denote the circumference of k by κ. Since in the resulting domain $\Delta\mathsf{K} = 0$, and the boundary integral over the boundary Γ vanishes, we are left with

$$\int_\kappa \left(u \frac{\partial \mathsf{K}}{\partial n} - \mathsf{K} \frac{\partial u}{\partial n} \right) ds = \iint_{G-k} \mathsf{K}\varphi(\xi, \eta)\, d\xi\, d\eta.$$

In the limit as $\epsilon \to 0$, $\int_\kappa u(\partial \mathsf{K}/\partial n)\, ds$ approaches $u(x, y)$ and $\int_\kappa \mathsf{K}(\partial u/\partial n)\, ds$ approaches zero, giving the desired result

$$u = \iint_G \mathsf{K}\varphi\, d\xi\, d\eta.$$

The second part of the theorem is proved most simply by means of an artifice introduced by Riemann, in which the assumed continuity of the first derivatives of $\varphi(x, y)$[1] is used. We decompose the function $u(x, y) = \iint_G \mathsf{K}(x, y; \xi, \eta)\varphi(\xi, \eta)\, d\xi\, d\eta$ into two terms corresponding to the decomposition $\mathsf{K} = -\log r/2\pi + \gamma(x, y; \xi, \eta)$ of Green's function, namely $u = \psi + \chi$, where

$$2\pi\psi(x, y) = -\iint_G \varphi(\xi, \eta) \log r\, d\xi\, d\eta,$$

$$\chi(x, y) = \iint_G \gamma(x, y; \xi, \eta)\varphi(\xi, \eta)\, d\xi\, d\eta.$$

Since the function $\gamma(x, y; \xi, \eta)$ and its derivatives up to the second order are everywhere continuous, we can at once form $\Delta\chi$ by differentiation under the integral sign and obtain $\Delta\chi = 0$ since $\Delta\gamma = 0$. Hence, to calculate Δu we need only compute $\Delta\psi$. Again, by differentiating under the integral sign we obtain the first derivative ψ_x.

[1] One cannot deduce the existence of continuous second derivatives of u merely from the continuity of φ. The assumption made in the text is, however, more stringent than necessary.

Introducing polar coordinates r, θ the integral $\iint_G \varphi(\xi, \eta) \log r \, d\xi \, d\eta$ assumes the form $\iint_G \varphi \, r \log r \, dr \, d\theta$; but if, before we introduce the polar coordinates, we differentiate with respect to x, the integral takes the form $\iint_G \varphi \cos\theta \, dr \, d\theta$, and the integrand remains continuous. If for the time being we set $-\log r/2\pi = S(x, y; \xi, \eta)$, we obtain

$$\psi_x = \iint_G S_x \varphi \, d\xi \, d\eta.$$

Since $S_x = -S_\xi$, we may also write

$$\psi_x = -\iint_G S_\xi \varphi \, d\xi \, d\eta.$$

Integrating this formula by parts, we may eliminate the derivative S_ξ and then once more differentiate under the integral sign. We obtain

$$\psi_x = -\int_\Gamma S\varphi \, d\eta + \iint_G S\varphi_\xi \, d\xi \, d\eta$$

and

$$\psi_{xx} = -\int_\Gamma S_x \varphi \, d\eta + \iint_G S_x \varphi_\xi \, d\xi \, d\eta = \int_\Gamma S_\xi \varphi \, d\eta - \iint_G S_\xi \varphi_\xi \, d\xi \, d\eta.$$

Similarly, we find

$$\psi_{yy} = -\int_\Gamma S_\eta \varphi \, d\xi - \iint_G S_\eta \varphi_\eta \, d\xi \, d\eta$$

and hence

$$\Delta\psi = \int_\Gamma \frac{\partial S}{\partial n} \varphi \, ds - \iint_G (S_\xi \varphi_\xi + S_\eta \varphi_\eta) \, d\xi \, d\eta.$$

If, on the right, instead of extending the double integral over the entire domain G we integrate over a domain G_ϵ obtained from G by cutting out a small circular disk k of radius ϵ and circumference κ about the point (x, y), then we can write

$$\Delta\psi = \int_\Gamma \frac{\partial S}{\partial n} \varphi \, ds - \lim_{\epsilon \to 0} \iint_{G_\epsilon} (S_\xi \varphi_\xi + S_\eta \varphi_\eta) \, d\xi \, d\eta.$$

In this expression we transform the double integral on the right according to Green's formula and find, since $\Delta S = 0$ everywhere in G,

$$\Delta\psi = \int_\Gamma \frac{\partial S}{\partial n}\varphi\, ds - \int_\Gamma \frac{\partial S}{\partial n}\varphi\, ds + \lim_{\epsilon\to 0}\int_\kappa \frac{\partial S}{\partial n}\varphi\, ds = \lim_{\epsilon\to 0}\int_\kappa \frac{\partial S}{\partial n}\varphi\, ds.$$

As we have already seen, the remaining boundary integral on the right tends to $-\varphi(x, y)$ as $\epsilon \to 0$. This proves that ψ satisfies "*Poisson's equation*" $\Delta f = -\varphi$.

For Poisson's equation in three dimensions, $\Delta u = -\varphi(x, y, z)$, and for the associated eigenvalue problem of the equation

$$\Delta u + \lambda u = 0,$$

we obtain results which correspond word for word to those for the two-dimensional case. In three dimensions, however, the singularity for Green's function is

$$\frac{1}{4\pi r} = \frac{1}{4\pi\sqrt{(x-\xi)^2 + (y-\eta)^2 + (z-\zeta)^2}},$$

and Green's function $K(x, y, z; \xi, \eta, \zeta)$ is of the form $K(x, y, z; \xi, \eta, \zeta) = 1/4\pi r + \gamma(x, y, z; \xi, \eta, \zeta)$, where $\gamma(x, y, z; \xi, \eta, \zeta)$ and its derivatives of first and second order are continuous. The function $1/4\pi r$ is itself a *fundamental solution of the equation* $\Delta u = 0$ (cf. pages 353 and 377).

In the case of partial differential equations, *the existence of Green's function* is far more difficult to establish than in the case of ordinary differential equations. General existence theorems will be proved in Volume II by means of the direct methods of the calculus of variations. At this point we must either postulate the existence of Green's function or limit ourselves to domains in which Green's function can be represented explicitly. In the next section we consider such domains. Once we have Green's function, however, the remaining discussion is parallel to that for ordinary differential equations. We consider here—for $\rho > 0$—the eigenvalue problem for the differential equation

$$\Delta v + \lambda\rho(x, y)v = 0 \tag{104}$$

under given homogeneous boundary conditions. As a consequence of the fundamental property of Green's function, we at once obtain from (104) the homogeneous integral equation

$$v(x, y) = \lambda \iint_G K(x, y; \xi, \eta)\rho(\xi, \eta)v(\xi, \eta)\, d\xi\, d\eta.$$

Let us introduce the symmetric kernel

$$K = \mathsf{K}\sqrt{\rho(x, y)\rho(\xi, \eta)}.$$

Evidently the function

$$u(x, y) = \sqrt{\rho(x, y)}v(x, y)$$

satisfies the symmetric homogeneous integral equation

$$u(x, y) = \lambda \iint_G K(x, y; \xi, \eta)u(\xi, \eta)\, d\xi\, d\eta; \tag{105}$$

since these relations may be inverted, the eigenvalue problem of equation (104) is seen to be completely equivalent with that of the symmetric integral equation (105). This integral equation may be treated by the theory of Chapter III; for, although the kernel becomes infinite at some point of the domain of integration, the integral $\iint_G \mathsf{K}(x, y; \xi, \eta)^2\, d\xi\, d\eta$ exists and is continuous in the variables x, y. *Hence we are assured of the existence of the eigenvalues* $\lambda_1, \lambda_2, \cdots$ *with* $\lambda_n \to \infty$ *for* $n \to \infty$ *and of an associated system of eigenfunctions* $v_1, v_2, \cdots$, *assumed to be normalized and denoted by* $u_1, u_2, \cdots$. According to the theorem on Green's function (cf. page 365) any function $w(x, y)$ which has continuous first and second derivatives and satisfies the boundary conditions may be represented by an integral of the form

$$w(x, y) = \iint_G K(x, y; \xi, \eta)h(\xi, \eta)\, d\xi\, d\eta$$

in terms of the function $h = -\Delta w$. Thus we obtain the result: *Every function* $w(x, y)$ *which satisfies the boundary conditions and possesses continuous derivatives up to the second order may be expanded in terms of the eigenfunctions in an absolutely and uniformly convergent series* $w = \sum_{n=1}^{\infty} c_n v_n(x, y)$ *with* $c_n = \iint_G \rho w v_n\, dx\, dy$. *Thus the normalized eigenfunctions* $\sqrt{\rho}\, v_n$ *form a complete orthogonal system.*

Since Green's function becomes infinite, Mercer's theorem cannot be applied here as it could in the case of ordinary differential equations. Therefore, in spite of the positive definite character of the kernel, we cannot conclude that the equation

$$\mathsf{K}(x, y; \xi, \eta) = \sum_{n=1}^{\infty} \frac{v_n(x, y)v_n(\xi, \eta)}{\lambda_n}$$

holds. Our general theory proves only the weaker relation

$$\lim_{m\to\infty} \iint_G \left[K - \sum_{n=1}^{m} \frac{v_n(x, y)v_n(\xi, \eta)}{\lambda_n} \right]^2 dx\, dy = 0.$$

The discussion of the general self-adjoint differential equation

$$p\Delta v + p_x v_x + p_y v_y - qv + \lambda\rho v = 0$$

is parallel to the previous discussion; it will therefore be sufficient to point out the fact that the results also remain unchanged. The only difference that should be mentioned is that now Green's function must have the form

$$K(x, y; \xi, \eta) = -\frac{a(x, y; \xi, \eta)}{2\pi p(\xi, \eta)} \log r + \gamma(x, y; \xi, \eta),$$

where $\gamma(x, y; \xi, \eta)$ and its derivatives up to the second order are continuous in a neighborhood of the singular point (although in general γ need no longer be a solution of the differential equation) and where a denotes a suitable function having continuous derivatives up to the second order, such that $a(\xi, \eta; \xi, \eta) = 1$ holds identically.

Similarly, for *partial differential equations of higher order*, the only essential difference lies in the form of the singularity associated with Green's function. If we consider, for example, the differential equation of a plate

$$\Delta\Delta v = -\varphi(x, y),$$

in two independent variables, then we must specify Green's function not only by boundary conditions and the requirement $\Delta\Delta K = 0$, but also by assuming that it has the form

$$K = -\frac{1}{8\pi} r^2 \log r + \gamma(x, y; \xi, \eta),$$

where $\gamma(x, y; \xi, \eta)$ and its derivatives up to the fourth order are continuous. The reader will easily verify that the given singularity is actually the correct one, i.e. that it corresponds to a point force. It should be stressed, moreover, that the function $r^2 \log r$ is itself a "fundamental solution" of $\Delta\Delta v = 0$.

In the case of a plate, passage to the corresponding integral equation again shows: *A sequence of eigenvalues and associated eigenfunctions which form a complete orthogonal system exists; every function which*

satisfies the boundary conditions and has continuous derivatives up to the fourth order in the fundamental domain G can be expanded in an absolutely and uniformly convergent series in terms of these eigenfunctions.

§15. *Examples of Green's Function*

1. Ordinary Differential Equations. Green's function for the expression

$$L[u] = u''$$

in the interval (0, 1) with the boundary conditions $u(0) = u(1) = 0$ is

$$\mathsf{K}(x, \xi) = \begin{cases} (1 - \xi)x & \text{for} \quad x \leq \xi, \\ (1 - x)\xi & \text{for} \quad x > \xi. \end{cases}$$

For the boundary conditions $u(0) = 0$, $u'(1) = 0$, Green's function becomes

$$\mathsf{K}(x, \xi) = \begin{cases} x & \text{for} \quad x \leq \xi, \\ \xi & \text{for} \quad x > \xi. \end{cases}$$

For the interval $-1 \leq x \leq +1$ and the boundary conditions

$$u(-1) = u(1) = 0$$

we find

$$\mathsf{K}(x, \xi) = -\tfrac{1}{2}\{|\, x - \xi \,| + x\xi - 1\},$$

an expression which could also be obtained from the first example by a transformation. On the other hand, in the interval $0 \leq x \leq 1$, for the boundary conditions $u(0) = -u(1)$, $u'(0) = -u'(1)$, we have

$$\mathsf{K}(x, \xi) = -\tfrac{1}{2}\,|\, x - \xi \,| + \tfrac{1}{4}.$$

Green's function for the differential expression

$$L[u] = xu'' + u'$$

associated with the Bessel function of zero-th order $J_0(x)$, for the interval $0 \leq x \leq 1$ and the boundary conditions $u(1) = 0$, $u(0)$ finite, has the form

$$\mathsf{K}(x, \xi) = \begin{cases} -\log \xi & \text{for} \quad x \leq \xi, \\ -\log x & \text{for} \quad x > \xi, \end{cases}$$

as may be easily ascertained by the general methods of the preceding section. Green's function, under the boundary conditions $u(1) = 0$, $u(0)$ finite, for the differential expression

$$L[u] = (xu')' - \frac{n^2}{x} u$$

associated with the Bessel function $J_n(x)$ (cf. eq. (28)), is given by

$$K(x, \xi) = \frac{1}{n}\left[\left(\frac{x}{\xi}\right)^n - (x\xi)^n\right] \qquad (x \leq \xi)$$

$$K(x, \xi) = \frac{1}{n}\left[\left(\frac{\xi}{x}\right)^n - (x\xi)^n\right] \qquad (x > \xi).$$

As a further example, consider the differential expression

$$L[u] = ((1 - x^2)u')' - \frac{h^2}{1 - x^2} u,$$

which for $h = 0, 1, 2, \cdots$ is associated with the Legendre functions of zero-th order, first order, etc., respectively; the interval of definition is $-1 \leq x \leq +1$, and the boundary conditions are: u finite at both end points. We can immediately specify solutions of $L[u] = 0$ which are finite at $x = -1$ and $x = 1$, namely, $c_1[(1 + x)/(1 - x)]^{h/2}$ and $c_2[(1 - x)/(1 + x)]^{h/2}$, respectively. Combining these solutions by the rules of §14, 2, we obtain Green's function:

$$K(x, \xi) = \frac{1}{2h}\left(\frac{1 + x}{1 - x}\frac{1 - \xi}{1 + \xi}\right)^{h/2} \qquad (x \leq \xi)$$

$$K(x, \xi) = \frac{1}{2h}\left(\frac{1 + \xi}{1 - \xi}\frac{1 - x}{1 + x}\right)^{h/2} \qquad (x > \xi).$$

According to the general theory, this construction fails for $h = 0$, since then the equation $L[u] = 0$ has the normalized solution $u = 1/\sqrt{2}$ which is everywhere regular and satisfies the boundary conditions. Thus for $h = 0$ it is necessary to find a Green's function in the generalized sense which satisfies the differential equation

$$L[u] = \tfrac{1}{2}.$$

This function is given by

$$K(x, \xi) = \begin{cases} -\frac{1}{2}\log[(1 - x)(1 + \xi)] + c & (x \leq \xi) \\ -\frac{1}{2}\log[(1 + x)(1 - \xi)] + c & (x > \xi), \end{cases}$$

where $c = \log 2 - \frac{1}{2}$.

Another simple example in which Green's function in the generalized sense occurs is the differential expression

$$L[u] = u''$$

for the interval $-1 \leq x \leq +1$ with the periodic boundary conditions $u(-1) = u(1)$, $u'(-1) = u'(1)$. Here again, there is a regular solution $u = 1/\sqrt{2}$ of $L[u] = 0$ which satisfies both boundary conditions (corresponding to the physical problem of a string with both ends free). Thus, we must construct Green's function in the generalized sense from the differential equation

$$u'' = \tfrac{1}{2}.$$

We easily obtain

$$\mathsf{K}(x, \xi) = -\tfrac{1}{2} \mid x - \xi \mid + \tfrac{1}{4}(x - \xi)^2 + \tfrac{1}{6}.$$

All these Green's functions serve as kernels of integral equations which correspond to the respective differential equation problems. We shall state explicitly the following bilinear formulas associated with our examples:

$$\frac{2}{\pi^2} \sum_{n=1}^{\infty} \frac{\sin n\pi x \sin n\pi \xi}{n^2} = \begin{cases} (1 - \xi)x & (x \leq \xi), \\ (1 - x)\xi & (x > \xi), \end{cases}$$

$$\frac{2}{\pi^2} \sum_{n=0}^{\infty} \frac{\sin (n + \frac{1}{2})\pi x \sin (n + \frac{1}{2})\pi \xi}{(n + \frac{1}{2})^2} = \begin{cases} x & (x \leq \xi), \\ \xi & (x > \xi), \end{cases}$$

furthermore

$$\mathsf{K}(x, \xi) = \sum_{n=1}^{\infty} \frac{(n + \frac{1}{2})P_n(x)P_n(\xi)}{n(n + 1)}$$

where

$$\mathsf{K}(x, \xi) = \begin{cases} -\frac{1}{2} \log [(1 - x)(1 + \xi)] + \log 2 - \frac{1}{2} & (x \leq \xi) \\ -\frac{1}{2} \log [(1 + x)(1 - \xi)] + \log 2 - \frac{1}{2} & (x > \xi). \end{cases}$$

Finally, we call particular attention to the Green's functions and integral equations associated with the Hermite and Laguerre polynomials and orthogonal functions.

The differential equation (49)

$$u'' + (1 - x^2)u + \lambda u = 0$$

of the orthogonal Hermite functions has the eigenvalue $\lambda = 0$ under the boundary conditions: regularity at plus and minus infinity. To avoid constructing Green's function in the generalized sense, consider the value $\lambda = -2$, which is certainly not an eigenvalue (cf. page 328); accordingly, we construct Green's function for the differential expression

$$L[u] = u'' - (1 + x^2)u$$

under the boundary condition that the solution vanish at $\pm\infty$. To obtain the general solution of the equation $L[u] = 0$, we note that $u(x) = e^{x^2/2}$ is a solution of $L[u] = 0$. Assuming the general solution to be of the form $u = we^{x^2/2}$, we at once obtain the equation

$$w'' + 2w'x = 0$$

for w, which in addition to the obvious solution $w = \text{const.}$ has also the solution

$$w = c_1 \int^x e^{-t^2}\, dt.$$

In this way, we find

$$u = c_1 e^{x^2/2} \int_{c_2}^x e^{-t^2}\, dt.$$

Hence the particular solutions which vanish at $x = +\infty$ and $x = -\infty$ are given by

$$ae^{x^2/2} \int_x^\infty e^{-t^2}\, dt \quad \text{and} \quad be^{x^2/2} \int_{-\infty}^x e^{-t^2}\, dt\,,$$

respectively. These lead at once to the following expression for Green's function:[1]

$$K(x, \xi) = \begin{cases} \dfrac{1}{\sqrt{\pi}}\, e^{(x^2+\xi^2)/2} \displaystyle\int_{-\infty}^x e^{-t^2}\, dt \int_\xi^\infty e^{-t^2}\, dt & (x \leq \xi) \\ \dfrac{1}{\sqrt{\pi}}\, e^{(x^2+\xi^2)/2} \displaystyle\int_{-\infty}^\xi e^{-t^2}\, dt \int_x^\infty e^{-t^2}\, dt & (x > \xi). \end{cases}$$

[1] Cf. R. Neumann, Die Entwicklung willkürlicher Funktionen etc., Dissertation, Breslau, 1912.

The factor $1/\sqrt{\pi}$ ensures the proper discontinuity of the derivative since the formula

$$\frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-t^2}\, dt = 1$$

holds.

The differential equation $L[u] + \lambda u = 0$ and the integral equation

$$u(x) = \lambda \int_{-\infty}^{\infty} K(x, \xi) u(\xi)\, d\xi$$

have the eigenvalues $\lambda = 2n+2 \quad (n = 0, 1, 2, \cdots)$ and the eigenfunctions

$$e^{-x^2/2} H_n(x).$$

The Laguerre functions $e^{-x/2} L_n(x)$ are solutions of the differential equation

$$xu'' + u' + \left(\frac{1}{2} - \frac{x}{4}\right) u + \lambda u = 0$$

for the eigenvalues $\lambda = n$ $(n = 0, 1, 2, \cdots)$. We consider this equation for the particular value $\lambda = -1$, and define

$$L[u] = xu'' + u' - \left(\frac{1}{2} + \frac{x}{4}\right) u.$$

The equation $L[u] = 0$ has the particular solution $e^{x/2}$. Assuming the general solution to be of the form

$$u = we^{x/2}$$

we obtain, as before, the expression

$$w = c_1 \int_{c_2}^{x} \frac{e^{-t}}{t}\, dt$$

for w, so that the two particular solutions which are regular at $x = 0$ and vanish for $x = +\infty$ are given by

$$ae^{x/2} \qquad \text{and} \qquad be^{x/2} \int_{x}^{\infty} \frac{e^{-t}}{t}\, dt,$$

respectively. Green's function for the boundary conditions stated in §10, 4, is constructed from these solutions:

$$K(x,\xi) = \begin{cases} e^{(x+\xi)/2}\int_{\xi}^{\infty}\frac{e^{-t}}{t}\,dt & \text{for} \quad x \leq \xi, \\ e^{(x+\xi)/2}\int_{x}^{\infty}\frac{e^{-t}}{t}\,dt & \text{for} \quad x > \xi. \end{cases}$$

No Green's function exists for the differential expression

$$L[u] = u''$$

in the interval $-\infty < x < \infty$, with the boundary condition: u finite. This corresponds to the fact that the homogeneous equation $u'' = 0$ has the solution $u = \text{const.}$ which is regular at infinity. On the other hand, the differential expression

$$L[u] = u'' - u$$

has the Green's function

$$\tfrac{1}{2}e^{-|x-\xi|},$$

and the singular integral equation

$$\varphi(x) = \frac{\lambda}{2}\int_{-\infty}^{\infty} e^{-|x-\xi|}\varphi(\xi)\,d\xi$$

obtained from this function has as its continuous spectrum all the values $\lambda = 1 + s^2 \geq 1$ with eigenfunctions $(\cos sx)/\sqrt{\pi}$, $(\sin sx)/\sqrt{\pi}$ (cf. §12). Here the bilinear relation is replaced by the integral formula

$$\frac{1}{\pi}\int_0^{\infty}\frac{\cos sx\cos s\xi + \sin sx\sin s\xi}{1+s^2}\,ds = \frac{1}{\pi}\int_0^{\infty}\frac{\cos s(x-\xi)}{1+s^2}\,ds = \frac{1}{2}e^{-|x-\xi|}.$$

As an example of Green's function for a fourth order differential expression, consider the equation $L[u] = u''''$ in the interval $0 \leq x \leq 1$, under the boundary conditions $u(0) = u(1) = u'(0) = u'(1) = 0$ (corresponding to a rod clamped at both ends). We find without difficulty

$$K(x,\xi) = \frac{x^2(\xi-1)^2}{6}(2x\xi + x - 3\xi) \quad \text{for} \quad x \leq \xi$$

and a corresponding expression for $x > \xi$.

2. Green's Function for Δu: Circle and Sphere. We consider the boundary condition, $u = 0$, and derive Green's function for the circle and sphere. We use the fact that the circle and the sphere are the geometrical loci of all points whose distances from two fixed points P_1, P_2 are in a constant ratio. More precisely, let $P_1:(\xi, \eta)$ or (ξ, η, ζ) be any point in the interior of the circle $x^2 + y^2 = 1$ or of the sphere $x^2 + y^2 + z^2 = 1$ and let P_2 be its reflection, having the coordinates $\xi/(\xi^2 + \eta^2)$, $\eta/(\xi^2 + \eta^2)$ (circle) or $\xi/(\xi^2 + \eta^2 + \zeta^2)$, $\eta/(\xi^2 + \eta^2 + \zeta^2)$, $\zeta/(\xi^2 + \eta^2 + \zeta^2)$ (sphere). If r_1, r_2 are the distances of the variable point $P:(x, y)$ or (x, y, z) from the points P_1, P_2, then the ratio $r_1:r_2$ is constant when the point P moves on the circumference of the circle or on the surface of the sphere, and the value of this ratio is given by $\sqrt{\xi^2 + \eta^2}$ or $\sqrt{\xi^2 + \eta^2 + \zeta^2}$. Now we note that the functions $-\log r_1/2\pi$, $-\log r_2/2\pi$ and $1/4\pi r_1$, $1/4\pi r_2$ are solutions of $\Delta u = 0$, and that $-\log r_1/2\pi$ and $1/4\pi r_1$, respectively, have the proper singularities for a fundamental solution at the point P_1. Thus the functions

$$K(x, y; \xi, \eta) = -\frac{1}{2\pi} \log \frac{r_1}{r_2} + \frac{1}{2\pi} \log \sqrt{\xi^2 + \eta^2}$$

and

$$K(x, y, z; \xi, \eta, \zeta) = \frac{1}{4\pi} \left(\frac{1}{r_1} - \frac{1}{r_2 \sqrt{\xi^2 + \eta^2 + \zeta^2}} \right)$$

are the Green's functions for the circle and sphere respectively under the boundary condition $u = 0$; for, these functions vanish on the respective boundaries.

3. Green's Function and Conformal Mapping. For two independent variables, it is helpful to use the function-theoretic fact that Green's function is related to the conformal map of the domain G onto the unit circle. Let $\zeta = f(x + iy)$ be an analytic function which maps the domain G conformally onto the unit circle in the ζ-plane in such a way that the point (ξ, η) of G goes into the origin. Then $-\log |f(x + iy)|/2\pi$ is the Green's function belonging to G. Thus we have Green's function for all domains which can be mapped conformally onto a circle. That these include all simply connected bounded domains with piecewise smooth boundaries is Riemann's fundamental theorem of geometric function theory.[1]

[1] Cf. Hurwitz-Courant, Funktionentheorie, pp. 389–423, in particular pp. 390–398, 3rd ed., J. Springer, Berlin, 1929.

4. Green's Function for the Potential Equation on the Surface of a Sphere. A simple example of Green's function in the generalized sense is given by the differential equation $\Delta^* u = 0$ (see §8 and §9, 1) with the condition of regularity at all points of the surface except the source point. Since the function $u = 1/\sqrt{4\pi}$ is regular on the whole surface, we must construct a Green's function in the generalized sense which satisfies the equation $\Delta^* u = 1/4\pi$. This function is obtained very easily by using the invariance of the expression $\Delta^* u$ under arbitrary rotations of the sphere. If we first put the source point P_1 of the Green's function at the north pole $\theta = 0$, we see that the differential equation $\Delta^* u = 1/4\pi$ is satisfied by the function $-\log [2 \sin (\theta/2)]/2\pi$ which depends only on the coordinate θ. Clearly, if $\rho(\theta, \varphi; \theta_1, \varphi_1)$ denotes the distance between two points $P:(\theta, \varphi)$ and $P_1:(\theta_1, \varphi_1)$ on the surface of the sphere, then from the invariance of $\Delta^* u$ under rotation it follows that

$$\mathsf{K}(\theta, \varphi; \theta_1, \varphi_1) = -\frac{1}{2\pi} \log \left(2 \sin \frac{\rho}{2}\right)$$

is a solution of $\Delta^* u = 1/4\pi$ which is regular at every point except $P = P_1$. Since this function, furthermore, has the proper singularity at $P = P_1$, it represents the desired Green's function. If we use K as the kernel of an integral equation

$$-2\pi Y(\theta, \varphi) = \lambda \iint_G \log \left(2 \sin \frac{\rho}{2}\right) Y(\theta_1, \varphi_1)\, d\theta_1\, d\varphi_1,$$

we find that the only eigenvalues and eigenfunctions associated with it are the $(2n + 1)$-fold eigenvalues $\lambda = n(n + 1)$ and the corresponding eigenfunctions $Y = Y_n(\theta, \varphi)$ defined in §9, 1.

5. Green's Function for $\Delta u = 0$ in a Rectangular Parallelepiped.[1] Let the boundary planes of the parallelepiped be given by $x = \pm a/2$, $y = \pm b/2$, $z = \pm c/2$. As a natural generalization of the method used in the case of the sphere (§15, 2), we shall find the Green's function for the boundary condition $u = 0$ by constructing the lattice which corresponds to the original parallelepiped and has the vertices $((k + \frac{1}{2})a, (m + \frac{1}{2})b, (n + \frac{1}{2})c)$ $(k, m, n = 0, \pm 1, \pm 2, \cdots)$ and reflecting the point (ξ, η, ζ) repeatedly across the lattice-planes. Thereby we obtain a system of points $(ka + (-1)^k \xi, mb + (-1)^m \eta,$

[1] We are obliged to A. Ostrowski for the convergence proof and the calculations of this section.

$nc + (-1)^n\zeta$). We imagine that, at each of these points, a unit of mass is concentrated which is positive if $k + m + n$ is even and negative if $k + m + n$ is odd. We would conjecture that the potential of such a mass distribution is equal to zero in the lattice-planes, since the contributions of the individual mass units cancel each other there. Thus we arrive at the following tentative expression for K:[1]

$$\mathsf{K} = \frac{1}{4\pi} \sum_{k=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \frac{(-1)^{k+m+n}}{\sqrt{N(k, m, n; \xi, \eta, \zeta; x, y, z)}} \tag{106}$$

where

$$\begin{aligned} N(k, m, n; \xi, \eta, \zeta; x, y, z) \\ = [ka + (-1)^k\xi - x]^2 + [mb + (-1)^m\eta - y]^2 \\ + [nc + (-1)^n\zeta - z]^2. \end{aligned}$$

Here, since the convergence can be at best conditional, we must first discuss the order of summation; to this end, we shall in general denote the expression $\varphi(k + 1) - \varphi(k)$, where $\varphi(k)$ is any function of k, by $\Delta_k\varphi(k)$. Then for fixed k and m, omitting the factor $(-1)^{k+m}$, we can write the inner sum with index n in the expression for K as

$$\begin{aligned} N'(k, m) &= \sum_{n=\pm1,\pm3,\cdots} \Delta_n \frac{1}{\sqrt{N(k, m, n)}} \\ &= -\sum_{n=0,\pm2,\pm4,\cdots} \Delta_n \frac{1}{\sqrt{N(k, m, n)}}, \end{aligned}$$

since $\lim_{|n|\to\infty} N(k, m, n) = \infty$. We apply the same transformation to the sums with indices m and k; since $\lim_{|m|\to\infty} N'(k, m) = 0$, as will be proved shortly, we have

$$N''(k) = \sum_{m=\pm1,\pm3,\cdots} \Delta_m N'(k, m) = -\sum_{m=0,\pm2,\pm4,\cdots} \Delta_m N'(k, m),$$

and also

$$\mathsf{K} = \frac{1}{4\pi} \sum_{k=\pm1,\pm3,\cdots} \Delta_k N''(k) = -\frac{1}{4\pi} \sum_{k=0,\pm2,\pm4,\cdots} \Delta_k N''(k),$$

because $\lim_{|k|\to\infty} N''(k) = 0$. Combining these results, we obtain the transformation

$$\mathsf{K} = \pm\frac{1}{4\pi} \sum_k \sum_m \sum_n \Delta_k \Delta_m \Delta_n \frac{1}{\sqrt{N(k, m, n)}}, \tag{107}$$

[1] Cf. Riemann and Hattendorf, Schwere, Elektrizität und Magnetismus, pp. 84–88, C. Rümpler, Hannover, 1880.

where each of the summation indices runs over either all the even or all the odd integers from $-\infty$ to $+\infty$. The entire sum has the sign $+$ if the summation runs over all the even integers an even number of times; it has the sign $-$ if the summation runs over them an odd number of times.

To prove our assertions it is sufficient to prove that the last sum converges absolutely. For this purpose we estimate its general term:

$$\left|\Delta_k \Delta_m \Delta_n \frac{1}{\sqrt{N(k, m, n)}}\right| < \frac{(d_1 |k| + c_1)(d_2 |m| + c_2)(d_3 |n| + c_3)}{(\sqrt{k^2 + m^2 + n^2})^7} < \frac{c}{(k^2 + m^2 + n^2)^2} \tag{108}$$

for

$$x^2 + y^2 + z^2 < h, \qquad \xi^2 + \eta^2 + \zeta^2 < h, \qquad k^2 + m^2 + n^2 > c_4(h),$$

$$d_1 = d_1(h), \cdots, \qquad c_3 = c_3(h), \qquad c = c(h).$$

This estimate is obtained if we apply the mean value theorem of the differential calculus three times and use the inequality relating the arithmetic and geometric means.

At the same time we also obtain uniform convergence in x, y, z, ξ, η, ζ if we sum only over k, m, n with $k^2 + m^2 + n^2 > c_4(h)$ such that $N(k, m, n)$ vanishes for none of the triples k, m, n.

For $x^2 + y^2 + z^2 < h$, $\xi^2 + \eta^2 + \zeta^2 < h$, $k^2 + m^2 + n^2 > c_4(h)$, the same consideration insures that all partial derivatives of the sum (107) obtained by termwise differentiation converge absolutely and that they converge uniformly in x, y, z, ξ, η, ζ.

It is now clear that (107) is the desired Green's function; naturally, (106) and (107) are meaningful only if no $N(k, m, n)$ vanishes. That conditions (1) and (3) (§14, 5) are fulfilled requires no proof. To show that condition (2) is satisfied, say in the plane $x = a/2$, we make use of the representation

$$\mathsf{K} = \frac{1}{4\pi} \sum_{k=\pm 1, \pm 3, \cdots} \Delta_k N''(k).$$

For $x = a/2$ the finite sums

$$\sum_{k=\pm 1, \pm 3, \cdots, \pm (l+1)} \Delta_k N''(k)$$

vanish because the individual terms cancel each other in pairs; hence we obtain $\mathsf{K} = 0$. In the same way it is seen that condition (2) is also fulfilled on the other planes of the parallelepiped.

The sum (106) was represented by Riemann as an integral over certain θ-products. This Riemann representation may be derived in the following manner: We write the equation

$$\frac{2}{\sqrt{\pi}} \int_0^\infty e^{-st^2}\, dt = \frac{1}{\sqrt{s}} \qquad (s > 0)$$

and substitute for s the expression $N(k, m, n; x, y, z; \xi, \eta, \zeta)$. This gives

$$\mathsf{K} = \frac{1}{2\pi\sqrt{\pi}} \sum_k \sum_m \sum_n \Delta_k \Delta_m \Delta_n \int_0^\infty e^{-Nt^2}\, dt.$$

If we could interchange summation and integration, we would find

$$\mathsf{K} = \frac{1}{2\pi\sqrt{\pi}} \int_0^\infty \sum_k \sum_m \sum_n \Delta_k \Delta_m \Delta_n\, e^{-Nt^2}\, dt$$

(109)

$$= \frac{1}{2\pi\sqrt{\pi}} \int_0^\infty f_1 f_2 f_3\, dt,$$

where the three factors under the integral sign are given by

$$f_1 = \sum_{k=-\infty}^{\infty} (-1)^k\, e^{-t^2[ka+(-1)^k\xi-x]^2},$$

$$f_2 = \sum_{m=-\infty}^{\infty} (-1)^m\, e^{-t^2[mb+(-1)^m\eta-y]^2},$$

$$f_3 = \sum_{n=-\infty}^{\infty} (-1)^n\, e^{-t^2[nc+(-1)^n\zeta-z]^2}$$

and may be expressed in terms of the theta function

$$\theta_{00}(z, \tau) = \theta_3(z, \tau) = \sum_{\nu=-\infty}^{\infty} e^{i\pi\nu^2\tau} e^{2i\pi\nu z}$$

We wish to prove formula (109); the main difficulty arises for $t = 0$, since the three series converge nonuniformly in its neighborhood. We prove first that we can interchange the order of integration and summation with respect to k:

$$\frac{1}{2\pi\sqrt{\pi}} \int_0^\infty f_1 f_2 f_3\, dt$$

(110)

$$= \frac{1}{2\pi\sqrt{\pi}} \sum_{k=-\infty}^{\infty} (-1)^k \int_0^\infty f_2 f_3\, e^{-t^2[ka+(-1)^k\xi-x]^2}\, dt.$$

It is easy to justify interchanging the order of summation and integration from 1 to ∞. In fact, for the remainder of the sum f_1 we have for $t > 1$, $p > P(\xi, x) > 2$ the estimate

$$\left| \sum_{|k|>p} (-1)^k e^{-t^2[k\alpha+(-1)^k\xi-x]^2} \right| < e^{-a^2t^2/4} \sum_{|k|>p} e^{-a^2t^2k^2/2}$$

$$< \frac{2e^{-a^2t^2/4}}{a^2} \sum_{|k|>p} \frac{1}{k^2} < \frac{2}{a^2} e^{-a^2t^2/4} \frac{1}{p-1} < \frac{4}{pa^2} e^{-a^2t^2/4};$$

thus the integral from 1 to ∞ converges to zero with increasing p. On the other hand, f_2 and f_3 clearly remain uniformly bounded in the interval from 1 to ∞.

To justify interchanging summation and integration from 0 to 1 it will be sufficient, according to a well-known theorem, to demonstrate the boundedness of the partial sums of the integrand. Now each of the two sums $\sum_{k=0}^{\infty}$ and $\sum_{k=1}^{\infty}$ into which f_1 may be decomposed is an alternating series whose terms, from a certain k on, decrease monotonically; k depends only on ξ and x, not on t. Hence the value of any partial sum of either series lies between fixed bounds for all $t > 0$. But a corresponding statement holds also for the partial sums of f_2 and f_3, which implies that f_2 and f_3 themselves are uniformly bounded for $t > 0$. Therefore we may apply the theorem referred to and prove equation (110). A precisely analogous consideration also proves that summation with respect to m and n may be interchanged with integration in the individual terms on the right-hand side of (109). This completes the proof of equation (109).

We shall now represent K in terms of the function θ_{00}. We have

$$f_1 = e^{-t^2(x-\xi)^2}\theta_{00}\left(-\frac{2at^2i(x-\xi)}{\pi}, \frac{4a^2t^2i}{\pi}\right)$$

$$-e^{-t^2(x+\xi)^2}\theta_{00}\left(-\frac{2at^2i(x+\xi)}{\pi}, \frac{4a^2t^2i}{\pi}\right),$$

$$f_2 = e^{-t^2(y-\eta)^2}\theta_{00}\left(-\frac{2bt^2i(y-\eta)}{\pi}, \frac{4b^2t^2i}{\pi}\right)$$

$$-e^{-t^2(y+\eta)^2}\theta_{00}\left(-\frac{2bt^2i(y+\eta)}{\pi}, \frac{4b^2t^2i}{\pi}\right),$$

$$f_3 = e^{-t^2(z-\zeta)^2}\theta_{00}\left(-\frac{2ct^2i(z-\zeta)}{\pi}, \frac{4c^2t^2i}{\pi}\right)$$

$$-e^{-t^2(z+\zeta)^2}\theta_{00}\left(-\frac{2ct^2i(z+\zeta)}{\pi}, \frac{4c^2t^2i}{\pi}\right).$$

To the individual terms we apply the transformation formula for the theta function:

$$\theta_{00}(z, \tau) = e^{-\pi i z^2/\tau} \frac{1}{\sqrt{-i\tau}} \theta_{00}\left(\frac{z}{\tau}, -\frac{1}{\tau}\right)$$

taking the principal value for the root. If in addition we set

$$q_x = e^{-\pi^2/4a^2t^2}, \qquad q_y = e^{-\pi^2/4b^2t^2}, \qquad q_z = e^{-\pi^2/4c^2t^2},$$

we find

$$\begin{aligned} f_1 &= \frac{\sqrt{\pi}}{2at}\left[\theta_{00}\left(-\frac{x-\xi}{2a}, \frac{\pi i}{4a^2t^2}\right) - \theta_{00}\left(-\frac{x+\xi}{2a}, \frac{\pi i}{4a^2t^2}\right)\right] \\ &= \frac{\sqrt{\pi}}{2at}\left[\sum_{k=-\infty}^{+\infty} q_x^{k^2} e^{-k(x-\xi)\pi i/a} - \sum_{k=-\infty}^{+\infty} q_x^{k^2} e^{-k(x+\xi)\pi i/a}\right] \\ &= \frac{\sqrt{\pi}}{at} \sum_{k=1}^{\infty} q_x^{k^2}\left(\cos\frac{k\pi(x-\xi)}{a} - \cos\frac{k\pi(x+\xi)}{a}\right) \\ &= \frac{2\sqrt{\pi}}{at} \sum_{k=1}^{\infty} q_x^{k^2} \sin\frac{k\pi x}{a} \sin\frac{k\pi\xi}{a}. \end{aligned} \tag{111}$$

Analogous expressions are obtained for f_2 and f_3, leading to the following expression for K:

$$\mathsf{K} = \frac{4}{abc}\int_0^{\infty} \frac{1}{t^3} \sum_{k=1}^{\infty}\sum_{m=1}^{\infty}\sum_{n=1}^{\infty} \sin\frac{k\pi x}{a} \sin\frac{k\pi\xi}{a} \cdots \sin\frac{n\pi\zeta}{c}\, e^{-(\pi^2/4t^2)[(k^2/a^2)+(m^2/b^2)+(n^2/c^2)]}\, dt.$$

Introducing $1/t^2 = \tau$ as a new variable of integration in this expression, we obtain

$$\mathsf{K} = \frac{2}{abc}\int_0^{\infty} \sum_{k=1}^{\infty}\sum_{m=1}^{\infty}\sum_{n=1}^{\infty} e^{-(\pi^2\tau/4)[(k^2/a^2)+(m^2/b^2)+(n^2/c^2)]} \cdot \sin\frac{k\pi x}{a} \cdots \sin\frac{n\pi\zeta}{c}\, d\tau.$$

This formula is a valid substitute for the expansion of Green's function in terms of the eigenfunctions

$$\mathsf{K}(x, y, z; \xi, \eta, \zeta) = \frac{8}{abc\pi^2} \sum_{k=1}^{\infty}\sum_{m=1}^{\infty}\sum_{n=1}^{\infty} \frac{\sin\frac{k\pi x}{a}\sin\frac{k\pi\xi}{a}\cdots\sin\frac{n\pi\zeta}{c}}{\frac{k^2}{a^2}+\frac{m^2}{b^2}+\frac{n^2}{c^2}},$$

which is obtained formally by interchanging summation and integration, but the convergence of which has not yet been proved.

As a simple expression for K, we find from (109), for $\tau = 1/t^2$,

$$\mathsf{K} = \frac{1}{32abc}\int_0^\infty \left\{\left[\theta_{00}\left(-\frac{x-\xi}{2a}, \frac{\pi i\tau}{4a^2}\right) - \theta_{00}\left(-\frac{x+\xi}{2a}, \frac{\pi i\tau}{4a^2}\right)\right]\cdots \right.$$
$$\left.\left[\theta_{00}\left(-\frac{z-\zeta}{2c}, \frac{\pi i\tau}{4c^2}\right) - \theta_{00}\left(-\frac{z+\zeta}{2c}, \frac{\pi i\tau}{4c^2}\right)\right]\right\} d\tau.$$

6. Green's Function for Δu in the Interior of a Rectangle. Consider the rectangle R with one vertex at the origin and the others at the points $(a, 0)$, $(0, b)$, (a, b). Let the source point be (ξ, η), and consider the point (x, y). If $\mathsf{K}(x, y; \xi, \eta)$ is the Green's function corresponding to the boundary condition $u = 0$, then K as a function of x and y is a solution of $\Delta u = 0$ in the interior of R, vanishes on the boundary, and becomes singular at the point (ξ, η) like $-\log r/2\pi$, where $r = \sqrt{(x-\xi)^2 + (y-\eta)^2}$. It seems reasonable, as in the case of a parallelepiped, to construct the lattice generated by the rectangle R, reflect the point (ξ, η) repeatedly across the lattice lines, and think of each of the resulting points as a source or a sink of intensity 1, depending on whether it is obtained from (ξ, η) by an even or by an odd number of reflections across the lattice lines.

As before, we could form the potential X of the resulting mass distribution by summing an infinite series. It is more convenient, however, to make use of function theory and introduce the associated analytic function $\varphi(x + iy) = X + iY$ whose real part is X. Then the function

$$f(x + iy) = e^{-2\pi(X+iY)} = e^{-2\pi\varphi(x+iy)}$$

must have simple zeros and poles at (ξ, η) and at the points resulting from reflections. Now take sets of four adjoining rectangles of our lattice and combine them to form rectangles of a new lattice. Then, in every rectangle of the new lattice, $f(x + iy)$ must have two simple zeros and two simple poles, which are symmetric with respect to the origin and congruent mod $(2a, 2b)$, respectively:

Zeros:	(ξ, η),	$(-\xi, -\eta)$,
Poles:	$(-\xi, \eta)$,	$(\xi, -\eta)$.

The simplest analytic function of this type is the elliptic function which has these zeros and poles in the period-rectangles with vertices

(a, b), $(-a, b)$, $(a, -b)$, $(-a, -b)$ and which may be represented in terms of the associated σ-function:

$$f(z) = \frac{\sigma(z - \xi - i\eta)\sigma(z + \xi + i\eta)}{\sigma(z - \xi + i\eta)\sigma(z + \xi - i\eta)},$$

where[1]

$$\sigma(z) = z \prod_{\omega}{}' \left[\left(1 - \frac{z}{2\omega}\right) e^{z/2\omega + z^2/8\omega^2)}\right],$$

$$\omega = ka + lbi \qquad (k = 0, \pm 1, \cdots) \qquad (l = 0, \pm 1, \cdots).$$

If we substitute this representation in the expression for $f(z)$ and multiply factorwise, we find, setting $\exp\{\xi\eta i/\omega^2\} = 1$ for $\omega = 0$,

$$f(z) = \prod_{\omega} \left[\frac{(z + \zeta - 2\omega)(z - \zeta - 2\omega)}{(z + \bar{\zeta} - 2\omega)(z - \bar{\zeta} - 2\omega)} e^{\xi\eta i/\omega^2}\right]$$

$$(\zeta = \xi + i\eta, \quad \bar{\zeta} = \xi - i\eta; \quad k = 0, \pm 1, \cdots, \quad l = 0, \pm 1, \cdots).$$

Here we have only to verify that the boundary condition is satisfied, i.e. that $f(z)$ has the absolute value 1 on the boundary of R. For $z = x = \mathcal{R}_e(z)$, the factor with $\omega = 0$ clearly has absolute value 1 and the factors corresponding to the remaining ω may be arranged in pairs corresponding to conjugate complex ω in such a way that the numerator of one factor is the complex conjugate of the denominator of the other factor of the pair. For $z = x + ib$ we multiply first with respect to l and then with respect to k. We can omit the factor $\exp\{\xi\eta i/\omega^2\}$ in the product taken with respect to l, since the sum $\sum 1/\omega^2$, taken over l for a fixed k, converges absolutely and has a real value. We combine the remaining factors pairwise in such a way that if one factor corresponds to $\omega = ka + lbi$, the second corresponds to $\omega = ka - (l - 1)bi$. It is then seen immediately that the product of such a pair has absolute value 1. However, for $z = iy$ we first take the product with respect to l and, for $|k| > 0$, combine any two such partial products which correspond to the values $\pm k$. We can then again omit the exponential factor, since $\sum 1/\omega^2$, taken over l, converges absolutely and has a real sum. We combine the remaining factors in pairs in such a way that one factor corresponds to $\omega = ka + lbi$, the other to $\omega = -ka + lbi$. Then every such product has absolute value 1. Finally, we take care of

[1] Π' denotes the product in which the factor corresponding to $\omega = 0$ is omitted.

the case $z = a + iy$ by combining factors which correspond to $\omega = ka + lbi$ and $\omega = -(k-1)a + lbi$, and by multiplying with respect to l. In this way we find for the desired Green's function the representation

$$\mathsf{K}(x, y; \xi, \eta) = -\frac{1}{2\pi}\,\mathcal{R}_e\left\{\log\frac{\sigma(z-\zeta, \omega_1, \omega_2)\sigma(z+\zeta, \omega_1, \omega_2)}{\sigma(z-\bar{\zeta}, \omega_1, \omega_2)\sigma(z+\bar{\zeta}, \omega_1, \omega_2)}\right\}$$

$$(z = x + iy, \quad \zeta = \xi + i\eta, \quad \bar{\zeta} = \xi - i\eta, \quad \omega_1 = a, \quad \omega_2 = ib).$$

The Green's function just constructed can be expanded in terms of the eigenfunctions $(2/\sqrt{ab}) \sin(k\pi x/a) \sin(m\pi y/b)$ into a convergent series.[1] The expansion is given by

$$\mathsf{K}(x, y; \xi, \eta) = \frac{4}{ab\pi^2}\sum_{m=1}^{\infty}\sum_{k=1}^{\infty}\frac{\sin k\frac{\pi}{a}x \sin m\frac{\pi}{b}y \sin k\frac{\pi}{a}\xi \sin m\frac{\pi}{b}\eta}{\frac{m^2}{b^2}+\frac{k^2}{a^2}}.$$

This is an example illustrating the validity of the bilinear formula, which has not been proved in general.

7. Green's Function for a Circular Ring. Consider a ring bounded by two circles with centers at the origin; the product of their radii is 1 (here the unit of length must be chosen in a suitable way). We denote the radius of the inner circle k_1 by $q^{1/2}$, and the radius of the outer circle k_2 by $q^{-1/2}$, where $0 < q < 1$. Then if c is the source point (which at first we assume to be real and positive), if $z = x + iy$ is the point under consideration, and if both points lie in the interior of the ring R, our problem is reduced to the following function-theoretic problem: Determine an analytic function $f(z)$ which has a simple zero at c, is analytic elsewhere in R, and has modulus 1 on the boundary of R. The desired Green's function is then to be obtained from $f(z)$ by

$$\mathsf{K}(x, y; \xi, \eta) = -\frac{1}{2\pi}\,\mathcal{R}_e\{\log f(z)\}.$$

In order to find enough function-theoretic properties to enable us to construct $f(z)$ explicitly, we now attempt to continue $f(z)$ beyond

[1] The convergence is not absolute, see V. A. Il'in, On the convergence of bilinear series of eigenfunctions, Uspehi Matem. Nauk (N. S.) 5, No. 4 (38), 1950, pp. 135–138.

the two circles. To this end, we associate to each point z in R a point z_1 inside k_1 so that $zz_1 = q$. If z moves toward the boundary of k_1, then the same is true of z_1, and in fact z_1 evidently approaches the complex conjugate point. Now, because of the symmetry of the assumptions, $f(z)$ may be considered a real function, that is, a function which takes real values at real points and, more generally, conjugate complex values at conjugate complex points. This implies that $f(z)f(q/z)$ approaches the positive real value $|f(z_0)|^2$ when z approaches a point z_0 on the circumference of the circle k_1. On the other hand, $f(z)$ has modulus 1 on the circumference of k_1. Thus for z on k_1 we have for $f(z)$ the equation

$$f(z)f\left(\frac{q}{z}\right) = 1, \tag{112}$$

and this equation holds identically for all z. Similarly, reflection across k_2 leads to the second functional equation

$$f(z)f\left(\frac{1}{qz}\right) = 1. \tag{113}$$

Since c is a simple zero of $f(z)$, it follows from successive application of these relations that $f(z)$ has simple zeros at the points

$$c, \quad q^{\pm 2}c, \quad q^{\pm 4}c, \quad \cdots$$

and simple poles at the points

$$q^{\pm 1}c^{-1}, \quad q^{\pm 3}c^{-1}, \quad q^{\pm 5}c^{-1}, \quad \cdots;$$

thus it agrees in its zeros and poles with the function

$$F(z) = \left(1 - \frac{z}{c}\right) \frac{\prod_{\nu=1}^{\infty} \left(1 - q^{2\nu}\frac{z}{c}\right)\left(1 - q^{2\nu}\frac{c}{z}\right)}{\prod_{\nu=1}^{\infty} (1 - q^{2\nu-1}cz)\left(1 - q^{2\nu-1}\frac{1}{cz}\right)}.$$

But for this function $F(z)$ we have the following functional equations of the type of (112) and (113):

$$F(z)F\left(\frac{q}{z}\right) = 1, \qquad F(z)F\left(\frac{1}{qz}\right) = \frac{1}{qc^2},$$

as can be verified by a simple calculation. Thus constants a and b can be determined in such a way that $az^bF(z)$ satisfies the functional

equations (112), (113) and has modulus 1 on the circles k_1, k_2, since a, b turn out to be real constants. We find the values

$$a = \pm\sqrt{c}\, q^{\frac{1}{4}}, \qquad b = -\tfrac{1}{2} - \frac{\log c}{\log q}.$$

Taking the negative sign for a, we obtain

$$f(z) = q^{\frac{1}{4}} z^{-\log c/\log q}\left(\sqrt{\frac{z}{c}} - \sqrt{\frac{c}{z}}\right)\frac{\prod_{\nu=1}^{\infty}\left(1 - q^{2\nu}\frac{z}{c}\right)\left(1 - q^{2\nu}\frac{c}{z}\right)}{\prod_{\nu=1}^{\infty}(1 - q^{2\nu-1}cz)\left(1 - q^{2\nu-1}\frac{1}{cz}\right)}$$

This expression may be written in terms of the theta functions

$$\theta_1(z) = -iCq^{\frac{1}{4}}(e^{i\pi z} - e^{-i\pi z})\prod_{\nu=1}^{\infty}(1 - q^{2\nu}e^{2i\pi z})(1 - q^{2\nu}e^{-2i\pi z}).$$

$$\theta_0(z) = C\prod_{\nu=1}^{\infty}(1 - q^{2\nu-1}e^{2i\pi z})(1 - q^{2\nu-1}e^{-2i\pi z})$$

with

$$C = \prod_{\nu=1}^{\infty}(1 - q^{2\nu}).$$

If we set $z = e^{2i\pi v}$, $c = e^{2i\pi\alpha}$, we find

$$f(z) = iz^{-2i\pi\alpha/\log q}\frac{\theta_1(v - \alpha)}{\theta_0(v + \alpha)},$$

and the real part of $\log f(z)$ naturally vanishes on k_1, k_2 also for complex c inside R; the problem is now solved.

§16. *Supplement to Chapter V*

1. Examples for the Vibrating String. (a) *The Plucked String.* We represent the solution for the case of a plucked string as a superposition of synchronous vibrations. At time $t = 0$ let the string be pulled a lateral distance h from the x-axis at the point $x = b$, so that the form of the string is linear from $x = 0$ to $x = b$ and from $x = b$ to $x = \pi$. Let the initial velocity be zero. Then the lateral displacement $u(x, t)$ has an expansion of the form

$$u(x, t) = \sum_{n=0}^{\infty} a_n \sin nx \cos nt,$$

where

$$a_n = \frac{2}{\pi}\int_0^\pi u(x, 0) \sin nx\, dx$$

$$= \frac{2h}{\pi}\left(\int_0^b \frac{x}{b} \sin nx\, dx + \int_b^\pi \frac{\pi - x}{\pi - b} \sin nx\, dx\right)$$

$$= \frac{2h}{n^2 b(\pi - b)} \sin nb.$$

Thus we find

$$u(x, t) = \frac{2h}{b(\pi - b)} \sum_{n=1}^{\infty} \frac{\sin nb \sin nx}{n^2} \cos nt.$$

(b) *Impulsive Excitation.* The case in which the vibration of the string is caused by an impulse applied to the string in its rest position in a neighborhood of the point $x = b$ can be treated in an analogous way. We obtain

$$u(x, t) = \sum_{n=1}^{\infty} b_n \sin nx \sin nt,$$

and

$$nb_n = \frac{2}{\pi}\int_0^\pi u_t(x, 0) \sin nx\, dx.$$

Now we must carry out a limit process, contracting the interval over which the impact acts, but doing this in such a way that the integral $\int_0^\pi u_t(x, 0)\, dx = U$ remains constant. In the limit we obtain:

$$b_n = 2U \frac{\sin nb}{\pi n},$$

$$u(x, t) = 2U \sum_{n=1}^{\infty} \frac{\sin nx \sin nb}{\pi n} \sin nt.$$

(c) *Forced Vibrations.* The general solution of the differential equation of forced vibration

$$u_{tt} - u_{xx} = f(x) \cos nt$$

with periodic external force has the form

$$u = -\frac{2}{\pi}\cos nt \sum_{\nu=1}^{\infty} \sin \nu x \frac{\int_0^\pi f(x) \sin \nu x \, dx}{n^2 - \nu^2} + \sum_{\nu=1}^{\infty} \sin \nu x (a_\nu \sin \nu t + b_\nu \cos \nu t).$$

If we let

$$\frac{2}{\pi}\int_0^\pi f(x) \sin \nu x \, dx = c_\nu,$$

we obtain for the corresponding integral, under the initial conditions $u(x, 0) = u_t(x, 0) = 0$, the expression

$$u(x, t) = -\sum_{\nu=1}^{\infty} \sin \nu x \frac{c_\nu}{n^2 - \nu^2}(\cos nt - \cos \nu t).$$

Here the term

$$\frac{-c_\nu}{n^2 - \nu^2} \sin \nu x \, (\cos nt - \cos \nu t)$$

generally dominates whenever n approaches the value ν. We can get the best idea of the behavior of this term by writing it in the form

$$\frac{2c_\nu}{n^2 - \nu^2} \sin \nu x \sin \frac{n + \nu}{2} t \sin \frac{n - \nu}{2} t;$$

we can think of this expression as the representation of a vibration $\sin [(n + \nu)t/2]$ with the variable amplitude $\sin [(n - \nu)t/2]$. The amplitude of the vibration becomes alternately strong and weak; this gives rise to the phenomenon of "*beats*." In the limit as $n \to \nu$, the term in question takes the form

$$\frac{c_\nu}{\nu} \sin \nu x \sin \nu t \cdot \frac{t}{2},$$

and thus the amplitude becomes infinite with time.

2. Vibrations of a Freely Suspended Rope; Bessel Functions. Let a homogeneous rope of length 1 and weight 1 hang along the x-axis, and let the direction of the weight be opposed to that of the x-axis.

We shall assume the rope is suspended at the point $x = 1$, so that the free end lies at the point $x = 0$. Then if u is the displacement perpendicular to the x-axis, u satisfies the differential equation.[1]

$$\frac{\partial^2 u}{\partial t^2} = \frac{\partial}{\partial x}\left(x \frac{\partial u}{\partial x}\right).$$

If we assume

$$u = q(t)\varphi(x),$$

we obtain the decomposition

$$\frac{\ddot{q}}{q} = -\lambda = \frac{(x\varphi')'}{\varphi}$$

with the auxiliary condition: $\varphi(1) = 0, \quad |\varphi(0)| < \infty$.

From this it follows that

$$\varphi(x) = cJ_0(2\sqrt{\lambda x}),$$

where $J_0(x)$ denotes the Bessel function of order 0 and the condition $J_0(2\nu) = 0$ determines a sequence of eigenfrequencies $\nu = \sqrt{\lambda}$.

3. Examples for the Explicit Solution of the Vibration Equation. Mathieu Functions. (a) *Sector of a Circle.* The vibration equation $\Delta u + \lambda u = 0$ for a circular segment $0 \leq r \leq 1, 0 \leq \theta \leq \alpha$ represented in polar coordinates is again solved by separation of variables: $u = f(r)g(\theta)$. Again, as in §9, we take $u = 0$ as boundary condition and find the system of eigenfunctions

$$u_n = \sin \frac{n\pi\theta}{\alpha} J_{n\pi/\alpha}(\sqrt{\lambda_{n,m}}r),$$

where $J_{n\pi/\alpha}$ denotes the Bessel function of index $n\pi/\alpha$ (cf. Ch. VII) and the eigenvalues $\lambda_{n,m}$ are determined by the transcendental equation $J_{n\pi/\alpha}(\sqrt{\lambda_{n,m}}) = 0$.

(b) *Ellipse.* The solution of the eigenvalue problem for the ellipse is obtained by introducing elliptic coordinates (see Ch. IV, §8, 3). We find

$$\Delta T + \lambda T = \frac{1}{\lambda_1 - \lambda_2}\left(\frac{\partial^2 T}{\partial t_1^2} - \frac{\partial^2 T}{\partial t_2^2}\right) + \lambda T = 0;$$

[1] Cf. Kneser, Integralgleichungen, pp. 39–43.

the assumption $T = U(t_1)V(t_2)$ leads to the equation

$$\frac{U''}{U} - \frac{V''}{V} = -\lambda(\lambda_1 - \lambda_2),$$

which is satisfied if and only if U and V are solutions either of the differential equations

$$U'' = -(\lambda\lambda_1 + \mu)U, \qquad V'' = -(\lambda\lambda_2 + \mu)V$$

or of

$$\frac{d^2U}{d\lambda_1^2} + \frac{1}{2}\left(\frac{1}{\lambda_1 - e_1} + \frac{1}{\lambda_1 - e_2}\right)\frac{dU}{d\lambda_1} = \frac{-\lambda\lambda_1 + \mu}{(\lambda_1 - e_1)(\lambda_1 - e_2)}\,U$$

and of the corresponding equation for V.

If we set

$$\frac{2\lambda_1 - e_1 - e_2}{e_1 - e_2} = \cosh u,$$

$$\frac{2\lambda_2 - e_1 - e_2}{e_1 - e_2} = \cos v,$$

then u and v are real. We obtain equations of the form

$$\frac{d^2U}{du^2} = -(\lambda' \cosh u + \mu')U,$$

$$\frac{d^2V}{dv^2} = (\lambda' \cos v + \mu')V,$$

where λ' and μ' are constants. The solutions of these differential equations, which are transformed into each other by the substitution $u = iv$, are called *functions of the elliptic cylinder* or *Mathieu functions*.[1]

(c) *Confocal Quadrilateral or Hexahedron.* Heretofore we have solved the vibration and potential equations by separation of variables for certain regions; these regions are either special cases or limiting cases of a quadrangle or a hexahedron bounded by members of a confocal system of cyclic curves or surfaces, respectively (cf. §9, 3).

4. Boundary Conditions with Parameters.[2] We shall describe briefly how certain boundary value problems with parameters in the bound-

[1] Cf. Whittaker and Watson, A Course of Modern Analysis, pp. 404–428.

[2] Cf. Hilbert, Integralgleichungen, pp. 77–81.

ary conditions can be reduced to integral equations. Consider, for example, the differential equation $\Delta u = 0$, with the following boundary condition on the regular boundary curve Γ of a simply connected domain G lying entirely in the finite part of the plane:

$$\frac{\partial u}{\partial n} + \lambda u + h(s) = 0,$$

where n denotes the outer normal, λ the parameter, and $h(s)$ a given function of arc length s on Γ. Using the Green's function $K(x, y; \xi, \eta)$ of the domain G whose normal derivatives vanish on the boundary, we obtain the Green's formula:

$$u(\xi, \eta) = \int_\Gamma [\lambda u(x, y) + h(s)]K(x, y; \xi, \eta)\, ds,$$

where the point (x, y) runs over the curve Γ. If we use the parametric representation $x = a(s)$, $y = b(s)$ for Γ, the values of $K(x, y; \xi, \eta)$ yield a symmetric function $K(s, \sigma)$ of two variables s, σ:

$$K(s, \sigma) = K(a(s), b(s); a(\sigma), b(\sigma)).$$

If in addition we let

$$u(a(s), b(s)) = \varphi(s),$$

$$\int_\Gamma K(s, \sigma)h(s)\, ds = f(\sigma),$$

then the above relation for u takes the form

$$f(\sigma) = \varphi(\sigma) - \lambda \int_\Gamma K(s, \sigma)\varphi(s)\, ds.$$

Only the solution of the first boundary value problem is needed to determine u from $\varphi(s)$; therefore, we have only to investigate this integral equation, whose kernel becomes logarithmically infinite at the single point $\sigma = s$. The general theory is immediately applicable to this kernel.

Analogous considerations hold for the general self-adjoint second order differential equation of the elliptic type.

5. Green's Tensors for Systems of Differential Equations. The idea which led us to introduce Green's function may be extended to problems involving systems of differential equations, for example, to the

problem of determining a vector $\mathbf{u}:(u_1, u_2, u_3)$ from a differential equation $L[\mathbf{u}] = -\mathbf{f}$, where $\mathbf{f}:(f_1, f_2, f_3)$ is a given vector. By a Green's tensor $\mathfrak{G}$ of the differential equation $L[\mathbf{u}] = -\mathbf{f}$ associated with prescribed homogeneous boundary conditions, for example $\mathbf{u} = 0$, we mean a matrix

$$\mathfrak{G}(x, y, z; \xi, \eta, \zeta) = \begin{pmatrix} K_{11} & K_{12} & K_{13} \\ K_{21} & K_{22} & K_{23} \\ K_{31} & K_{32} & K_{33} \end{pmatrix}$$

with the following properties: the differential equation $L[\mathbf{u}] = -\mathbf{f}$ together with the boundary conditions is equivalent to the formula

$$\mathbf{u}(x, y, z) = \iiint \mathfrak{G}(x, y, z; \xi, \eta, \zeta)\mathbf{f}(\xi, \eta, \zeta)\, d\xi\, d\eta\, d\zeta\ .$$

Here $\mathfrak{G}\mathbf{f}$ denotes the vector resulting from matrix multiplication of the matrix $\mathfrak{G}$ with the vector $\mathbf{f}$, that is, the vector with the components

$$K_{11}f_1 + K_{12}f_2 + K_{13}f_3\ ,\quad K_{21}f_1 + K_{22}f_2 + K_{23}f_3\ ,\quad K_{31}f_1 + K_{22}f_2 + K_{33}f_3.$$

Each column of the Green's tensor represents a vector $\mathbf{k}_i$, which, except at the *source point* $x = \xi, y = \eta, z = \zeta$, is continuous together with its derivatives and satisfies the differential equation $L[\mathbf{k}_i] = 0$ and the boundary conditions. The nature of its singularity at the source point becomes clear from its interpretation (e.g. in the case of a single differential equation) as the influence function of a point force acting at the source point $x = \xi, y = \eta, z = \zeta$. Green's tensor satisfies the symmetry relations

$$K_{ii}(x, y, z; \xi, \eta, \zeta) = K_{ii}(\xi, \eta, \zeta; x, y, z),$$

$$K_{ik}(x, y, z; \xi, \eta, \zeta) = K_{ki}(\xi, \eta, \zeta; x, y, z),$$

whenever (as we shall assume) the differential expression $L[\mathbf{u}]$ is self-adjoint, i.e. results from the variation of a quadratic differential expression in the vector $\mathbf{u}$ and its first derivatives. The eigenvalue problem for the differential equation $L[\mathbf{u}] + \lambda\mathbf{u} = 0$ may be solved with the aid of Green's tensor in a manner completely analogous to the ordinary case.[1]

[1] Cf. Hilbert, Integralgleichungen, pp. 206–212.

6. Analytic Continuation of the Solutions of the Equation $\Delta u + \lambda u = 0$. Let a solution of $\Delta u + \lambda u = 0$ be given in a closed domain G, which has the line-segment l as part of its boundary. Let this solution be continuous together with its derivatives up to the second order and let the function u or the normal derivative $\partial u/\partial n$ vanish on l. We can reflect G across l, obtaining a new domain G', and continue the function u into G' in the following way: If g' is the mirror-image of the point g of G under reflection, let $\bar{u}(g') = -u(g)$ when $u = 0$ on l and $\bar{u}(g') = u(g)$ when $\partial u/\partial n = 0$ on l. Then $\bar{u}$ is an extension of u which is a continuous solution of $\Delta u + \lambda u = 0$ in the combined domain $G + G'$ with continuous derivatives up to the second order.[1] Similar theorems may be stated for the equation $\Delta\Delta u - \lambda u = 0$ of a plate. The assumption for the theorem may be further weakened, as in the case of the reflection principle in function theory; this will be indicated later.

7. A Theorem on the Nodal Curves of the Solutions of $\Delta u + \lambda u = 0$. *If several branches of the curve $u = 0$ intersect in the interior of a domain of the x, y-plane in which u is regular,*[2] *then the set of nodal lines which meet at the point of intersection forms an equiangular system of rays.* The reader may prove this theorem by expanding the function u at the point in question in a power series.

8. An Example of Eigenvalues of Infinite Multiplicity. For an arbitrary plane domain, e.g. a circle, we consider the eigenvalue problem of $\Delta\Delta u - \lambda u = 0$ with the boundary conditions $\Delta u = 0$, $(\partial/\partial n)\Delta u = 0$. We easily obtain infinitely many eigenvalues λ_h and eigenfunctions u_h for this problem by noting that the functions $\Delta u_h = v_h$ must be eigenfunctions for the clamped plate, as long as Δu_h does not vanish identically. Thus we are led to eigenvalues which coincide with those for the clamped plate; zero also occurs as an eigenvalue of infinite multiplicity. Namely, for $\lambda = 0$ each of the infinitely many linearly independent potential functions, regular in G, satisfies the equation $\Delta\Delta u + \lambda u = 0$ under the given boundary conditions.

9. Limits for the Validity of the Expansion Theorems. In stating

[1] Cf. R. Courant, Beweis des Satzes etc., Math. Zeitschr., Vol. 1, 1918, pp. 321-328.

[2] It is not difficult to see that every continuous solution u with continuous derivatives is a regular analytic function of x and y (cf. also Vol. II).

our theorems on expansion in terms of the eigenfunctions of the differential equation

$$L[u] + \lambda\rho u = 0,$$

we have always made the assumption $\rho > 0$. That this assumption is essential is shown by the following example: In the equation $y'' + \lambda\rho y = 0$, let $\rho = 0$ for an arbitrary subinterval of the fundamental domain. Then every eigenfunction must be linear in this subinterval; hence the expansion theorem cannot hold for "arbitrary" functions.

References

Bôcher, M., Über die Reihenentwicklungen der Potentialtheorie. B. G. Teubner, Leipzig, 1894.

——, Leçons sur les méthodes de Sturm dans la théorie des équations différentielles linéaires et leurs développements modernes. Gauthier-Villars, Paris, 1917.

Courant, R., Zur Theorie der kleinen Schwingungen. Zeitschr. f. angew. Math. u. Mech., Vol. 2, 1922, pp. 278–285.

Frank, Ph., and v. Mises, R., Die partiellen Differential- und Integralgleichungen der Mechanik und Physik. F. Vieweg und Sohn, Braunschweig, 1925, 1927.

Hilbert, D., Grundzüge einer allgemeinen Theorie der linearen Integralgleichungen. B. G. Teubner, Leipzig and Berlin, 1912. (Cited as "Integralgleichungen.")

Hort, W., Technische Schwingungslehre. 2nd ed., J. Springer, Berlin, 1922.

Kneser, A., Die Integralgleichungen und ihre Anwendungen in der mathematischen Physik. 2nd ed., F. Vieweg und Sohn, Braunschweig, 1922. (Cited as "Integralgleichungen.")

Pockels, F., Über die partielle Differentialgleichung $\Delta u + k^2 u = 0$ und deren Auftreten in der mathematischen Physik. B. G. Teubner, Leipzig, 1891.

Rayleigh, J. W. S., The Theory of Sound. 2 vols., 2nd ed., Macmillan and Co., London, 1894, 1896.

Riemann, B., and Hattendorf, K., Schwere, Elektrizität und Magnetismus. C. Rümpler, Hannover, 1880.

Weber, H., Die partiellen Differentialgleichungen der mathematischen Physik. 2 vols., 4th ed., F. Vieweg und Sohn, Braunschweig, 1900, 1901; 5th ed., 1910, 1912.

Whittaker, E. T., and Watson, G. N., A Course of Modern Analysis. 4th ed., Cambridge University Press, Cambridge, 1927.

CHAPTER VI

Application of the Calculus of Variations to Eigenvalue Problems

In the preceding chapter we saw that eigenvalue problems of differential equations are closely connected with those of quadratic forms. As a matter of fact, these eigenvalue problems correspond to the problem of transforming a quadratic form in infinitely many variables to principal axes.

If, for example, we denote the potential and kinetic energies of a one-dimensional continuum by

$$U = \frac{1}{2}\int_0^\pi p\left(\frac{\partial u}{\partial x}\right)^2 dx \quad \text{and} \quad T = \frac{1}{2}\int_0^\pi \rho\left(\frac{\partial u}{\partial t}\right)^2 dx,$$

we need only write $u = \sum_{\nu=1}^{\infty} f_\nu(t) \sin \nu x$ and expand p and ρ in Fourier series; then we can consider the two expressions U and T as quadratic forms in the infinitely many variables (coordinates) f_ν and $\dot{f}_\nu$, respectively. If an orthogonal transformation

$$f_\nu = \sum_{\mu=1}^{\infty} t_{\nu\mu} q_\mu, \qquad \dot{f}_\nu = \sum_{\mu=1}^{\infty} t_{\nu\mu} \dot{q}_\mu \qquad (\nu = 1, 2, \cdots)$$

of these variables into new variables q_ν and $\dot{q}_\nu$ can be determined in such a way that U and T take the form

$$U = \sum_{\nu=1}^{\infty} \lambda_\nu q_\nu^2, \qquad T = \sum_{\nu=1}^{\infty} \dot{q}_\nu^2,$$

then the numbers λ_ν are the eigenvalues of our vibration problem. Since the eigenvalues of a quadratic form are characterized by simple extremum properties, it seems reasonable to consider a similar characterization for the case of quadratic functionals. However, instead of treating these functionals as limits of finite quadratic forms of a finite number of variables, we shall formulate and apply the extremum properties in question directly without referring to such a limit process.

§1. *Extremum Properties of Eigenvalues*

1. Classical Extremum Properties. Consider the eigenvalue problem of a self-adjoint second order partial differential equation

$$L[u] + \lambda\rho u = (pu_x)_x + (pu_y)_y - qu + \lambda\rho u = 0 \tag{1}$$

$$(p > 0,\ \rho > 0);$$

here x and y are independent variables in the fundamental domain G, bounded by one or several continuous curves Γ with piecewise continuous tangents. Let the boundary condition have the form $u = 0$, or in general $\partial u/\partial n + \sigma u = 0$, where σ denotes a piecewise continuous function of position on the boundary Γ and $\partial/\partial n$ differentiation in the direction of the outer normal.[1] Typical quadratic functional expressions for the variational eigenvalue problems are:

$$\mathfrak{D}[\varphi] = D[\varphi] + \int_\Gamma p\sigma\varphi^2\, ds \tag{2}$$

with

$$D[\varphi] = \iint_G p(\varphi_x^2 + \varphi_y^2)\, dx\, dy + \iint_G q\varphi^2\, dx\, dy \tag{2a}$$

and

$$H[\varphi] = \iint_G \rho\varphi^2\, dx\, dy; \tag{3}$$

the associated polar forms

$$\mathfrak{D}[\varphi, \psi] = D[\varphi, \psi] + \int_\Gamma p\sigma\varphi\psi\, ds,$$

$$D[\varphi, \psi] = \iint_G p(\varphi_x\psi_x + \varphi_y\psi_y)\, dx\, dy + \iint_G q\varphi\psi\, dx\, dy,$$

$$H[\varphi, \psi] = \iint_G \rho\varphi\psi\, dx\, dy$$

[1] It should be stated that in general this boundary condition must be interpreted in the "weak" sense; i.e., the function is not actually required to assume boundary values everywhere on Γ. This delicate point will be fully discussed in Vol. II, Ch. VII in connection with the existence proof. Here, where we assume the existence of the solutions, we need not specify the behavior of u at the boundary.

satisfy the relations

$$\mathfrak{D}[\varphi + \psi] = \mathfrak{D}[\varphi] + 2\mathfrak{D}[\varphi, \psi] + \mathfrak{D}[\psi],$$

$$H[\varphi + \psi] = H[\varphi] + 2H[\varphi, \psi] + H[\psi].$$

We require the argument function φ to be continuous in $G + \Gamma$ and to have piecewise continuous first derivatives in G.

Here piecewise continuity of a function ψ in G means: G may be subdivided into a finite number of regions $G_1, G_2, \cdots, G_m$ by arcs with piecewise continuous tangents such that ψ is continuous in each closed subregion of G not containing any of these arcs in its interior.

If the boundary condition $u = 0$ is prescribed, then u is assumed to be continuous in the closed region $G + \Gamma$ and to vanish on Γ.

We obtain the eigenvalues λ_ν and the associated eigenfunctions u_ν of the differential equation (1) from the following minimum properties: *The admissible function which minimizes the expression* $\mathfrak{D}[\varphi]$ *under the condition* $H[\varphi] = 1$ *is an eigenfunction* u_1 *for the differential equation* (1) *and satisfies*[1] *the natural boundary condition* $\partial\varphi/\partial n + \sigma\varphi = 0$; *the minimum value of* $\mathfrak{D}$ *is the corresponding eigenvalue. If we impose not only the condition*

$$H[\varphi] = 1 \tag{3a}$$

but also the orthogonality condition

$$H[\varphi, u_1] = 0,$$

then the solution is again an eigenfunction u_2 *of* (1) *satisfying the same boundary condition, and the minimum value* $\mathfrak{D}[u_2] = \lambda_2$ *is the associated eigenvalue. The successive minimum problems,* $\mathfrak{D}[\varphi] =$ *minimum subject to the condition* $H[\varphi] = 1$ *and to the auxiliary conditions*

$$H[\varphi, u_i] = 0 \qquad (i = 1, 2, \cdots, \nu - 1),$$

define the eigenfunctions u_ν *of equation* (1) *with the boundary condition* $\partial\varphi/\partial n + \sigma\varphi = 0$; *the associated eigenvalue* λ_ν *equals the minimum value* $\mathfrak{D}[u_\nu]$.

We can omit the normalizing condition $H[\varphi] = 1$ by minimizing the quotients $\mathfrak{D}[\varphi]/H[\varphi]$; then the solution is determined only up to an arbitrary factor of proportionality.

If the boundary condition is $u = 0$, the same variational problems

[1] See footnote 1 on previous page.

define eigenvalues and eigenfunctions; but the boundary condition $\varphi = 0$ is now added to the admissibility conditions. Then the boundary term $\int_\Gamma p\sigma\varphi^2\,ds$ in $\mathfrak{D}[\varphi]$ drops out automatically.

A proof that our minimum problem does possess solutions with continuous second derivatives will be given in Volume II, in connection with the direct methods of the calculus of variations. At present we proceed on the assumption that minimum problems of the above type possess solutions.

We have to show (a) that the solutions of the variational problem are also eigenfunctions for our differential equation problem, and (b) that they furnish *all* the eigenfunctions. The second assertion will be proved in §3 by showing that the system of functions $u_1, u_2, \cdots$ obtained from the variational problem is complete. Statement (a) could be derived from the general multiplier rule of Ch. IV, §7 but will now be obtained directly.

Let us begin with the first variational problem, and assume that its solution u_1 satisfies the condition $H[u_1] = 1$. If ζ is any function which satisfies the same conditions as φ, and if ϵ is an arbitrary constant, then for every value of ϵ and for $u = u_1, \lambda = \lambda_1$ we have

$$\mathfrak{D}[u + \epsilon\zeta] \geq \lambda H[u + \epsilon\zeta]$$

or

$$2\epsilon\left\{\mathfrak{D}[u, \zeta] - \lambda H[u, \zeta] + \frac{\epsilon}{2}(\mathfrak{D}[\zeta] - \lambda H[\zeta])\right\} \geq 0;$$

these statements are equivalent in view of the relation $\mathfrak{D}[u] = \lambda H[u]$. The second inequality can be valid for arbitrary values of ϵ only if equation

$$\mathfrak{D}[u, \zeta] - \lambda H[u, \zeta] = 0 \tag{4}$$

holds, i.e. if the first variation of the expression $\mathfrak{D} - \lambda H$ vanishes.

We now transform the expression $\mathfrak{D}[u, \zeta]$ according to Green's formula

$$\mathfrak{D}[u, \zeta] = -\iint_a \zeta L[u]\,dx\,dy + \int_\Gamma p\sigma\zeta u\,ds + \int_\Gamma p\zeta\frac{\partial u}{\partial n}\,ds$$

(cf. Ch. V, §1); since the function ζ is arbitrary, we immediately obtain equation (1) for $u = u_1$ and $\lambda = \lambda_1$. Considering the second

minimum problem with the additional condition $H[\varphi, u_1] = 0$, we obtain equation (4) for $u = u_2$ and $\lambda = \lambda_2$ at first only under the assumption that ζ satisfies relation

$$H[\zeta, u_1] = 0. \tag{5}$$

Now if η is any function having piecewise continuous derivatives of second order, we can determine a number t in such a way that the function $\zeta = \eta + tu_1$ satisfies condition (5)—we set $t = -H[u_1, \eta]$. Moreover, we note that we can, in particular, substitute the function $\zeta = u_2$ in equation (4) with $u = u_1$ and $\lambda = \lambda_1$; since u_2 satisfies the additional condition

$$H[u_2, u_1] = 0 \tag{6}$$

we immediately obtain

$$\mathfrak{D}[u_2, u_1] = 0. \tag{7}$$

If we substitute our function $\zeta = \eta + tu_1$ in equation (4), writing $u = u_2, \lambda = \lambda_2$, we find

$$\mathfrak{D}[u, \eta] - \lambda H[u, \eta] + t(\mathfrak{D}[u, u_1] - \lambda H[u, u_1]) = 0$$

or, taking into consideration equations (6) and (7),

$$\mathfrak{D}[u, \eta] - \lambda H[u, \eta] = 0. \tag{4a}$$

In other words, equation (4) holds also for arbitrary functions η or ζ without regard to the auxiliary condition (5). From this it follows directly, as above, that equation (1) is valid for $u = u_2$ and $\lambda = \lambda_2$. Continuing in this way, we conclude that the eigenvalue equation (4a) holds generally for the solutions u_i and the minimum values λ_i. For the solutions of the problem, which are normalized in accordance with (3a), we have the relations

$$\begin{aligned} &\mathfrak{D}[u_i] = \lambda_i, \qquad \mathfrak{D}[u_i, u_k] = 0 \\ &H[u_i] = 1, \qquad H[u_i, u_k] = 0 \end{aligned} \qquad (i \neq k). \tag{8}$$

The eigenvalues thus obtained satisfy the inequality

$$\lambda_{\nu-1} \leq \lambda_\nu \tag{9}$$

since the class of functions φ admissible in the problem characterizing λ_ν is a subset of the corresponding class admissible in the previ-

ous problem characterizing $\lambda_{\nu-1}$. Thus the minimum λ_ν can be no smaller than the preceding minimum $\lambda_{\nu-1}$.

Our variational problems produce an infinite sequence of eigenvalues and eigenfunctions of the associated differential equation problem. Conversely, the eigenvalues and eigenfunctions associated with the differential equation are obtained as solutions of our variational eigenvalue problems, as will be shown in §3, 1.

2. Generalizations.[1] It need hardly be pointed out that the other eigenvalue problems considered in the preceding chapter can also be treated from the standpoint of the calculus of variations. This is equally true of single and multiple integrals and of differential equations of second and higher order. For example, the eigenvalue problem of the Sturm-Liouville equation

$$(pu')' - qu + \lambda\rho u = 0$$

with the boundary conditions $u'(0) - h_1u(0) = u'(\pi) + h_2u(\pi) = 0$ corresponds to a variational problem of the form

$$\mathfrak{D}[\varphi] = \int_0^\pi (p\varphi'^2 + q\varphi^2)\,dx + h_1p(0)\varphi(0)^2 + h_2p(\pi)\varphi(\pi)^2 = \min.$$

without boundary conditions. All the homogeneous boundary conditions previously listed can be obtained by an appropriate choice of h_1 and h_2, if we consider the boundary conditions $u(0) = u(\pi) = 0$ as limiting cases for infinite h_1 and h_2.

Even for Sturm-Liouville equations with singularities at end points, the eigenvalues and eigenfunctions may be characterized by variational problems. It will suffice to give the formulation for Legendre polynomials and for Bessel functions. We obtain the Legendre polynomials by considering the free problem with

$$\mathfrak{D}[\varphi] = \int_{-1}^{+1} (1 - x^2)\varphi'^2\,dx, \qquad H[\varphi] = \int_{-1}^{+1} \varphi^2\,dx;$$

here both end points are singular. The Bessel function $J_0(x\sqrt{\lambda})$ of order zero arises from the problem for which

$$\mathfrak{D}[\varphi] = \int_0^1 x\varphi'^2\,dx, \qquad H[\varphi] = \int_0^1 x\varphi^2\,dx$$

[1] See Courant, Über die Anwendung der Variationsrechnung. . . .

with no boundary condition at $x = 0$, the Bessel functions of m-th order with $m \geq 1$ from the problem for which

$$\mathfrak{D}[\varphi] = \int_0^1 \left(x\varphi'^2 + \frac{m^2}{x}\varphi^2 \right) dx, \qquad H[\varphi] = \int_0^1 x\varphi^2 \, dx$$

with the boundary condition $\varphi(0) = 0$. The boundary condition $\varphi(1) = 0$ at the regular end point $x = 1$ is associated with the problem of a clamped membrane.

Analogous results are obtained for self-adjoint differential equations of higher order in more dimensions, such as the equation of the vibrating plate,

$$\Delta\Delta u - \lambda u = 0. \tag{10}$$

To treat the problem of a clamped plate (cf. Ch. IV, §10) we set

$$\mathfrak{D}[\varphi] = D[\varphi] = \iint_G (\Delta\varphi)^2 \, dx \, dy, \qquad H[\varphi] = \iint_G \varphi^2 \, dx \, dy$$

and impose the condition

$$\varphi = \frac{\partial\varphi}{\partial n} = 0$$

on the boundary Γ of G. With these conventions the rest of the discussion and the formulas of subsection 1 remain entirely unchanged.

Other types of eigenvalue problems, not treated explicitly in Chapter V, immediately fit into the scheme presented here. If we recall that $\frac{1}{2}H[\varphi]$ corresponds to the kinetic energy of our continuum with mass density ρ, while $\frac{1}{2}\mathfrak{D}[\varphi]$ represents the potential energy, it is natural to construct mechanical models in which the mass is not only distributed continuously over the domain G but concentrated at isolated points. For models of this kind we obtain (in one dimension) the expression

$$\mathfrak{H}[\varphi] = \int_G \rho\varphi^2 \, dx + \sum_{\nu=1}^{h} b_\nu \varphi(x_\nu)^2 \tag{11}$$

which takes the place of H; here $x_1, x_2, \cdots, x_h$ denote given points in the domain G and the numbers b_ν are given constants. A quadratic functional of this form corresponds to the assumption that masses of magnitude b_ν are concentrated at the points $x_1, x_2, \cdots, x_h$. We shall assume throughout that these masses are non-nega-

tive. Similarly, we can consider more general expressions of the form

$$\mathfrak{T}[\varphi] = \int_G p\varphi'^2\,dx + \int_G q\varphi^2\,dx + \sum_{\nu=1}^{h} a_\nu\varphi(x_\nu)^2. \tag{12}$$

For such problems we find eigenvalues and eigenfunctions using precisely the same notation and argument as in subsection 1. These eigenfunctions satisfy the differential equation

$$L[u] + \lambda\rho u = (pu')' - qu + \lambda\rho u = 0 \tag{13}$$

except at the points $x_1, x_2, \cdots, x_h$, where natural boundary and *jump conditions* for the derivatives occur. We obtain these conditions immediately by forming the first variation. The eigenfunctions of our problem, multiplied by $\sqrt{\rho}$, are no longer orthogonal; instead, they satisfy the conditions[1]

$$\int_G \rho u_i u_j\,dx + \sum_{\nu=1}^{h} b_\nu u_i(x_\nu)u_j(x_\nu) = \begin{cases} 0 & \text{for} \quad i \neq j, \\ 1 & \text{for} \quad i = j. \end{cases} \tag{14}$$

A further example is presented by the expressions

$$\mathfrak{D}[\varphi] = \int_G p\varphi'^2\,dx + \int_G q\varphi^2\,dx \tag{15}$$

and

$$\mathfrak{H}[\varphi] = \int_G \rho\varphi^2\,dx + \int_G\int_G k(x, y)\varphi(x)\varphi(y)\,dx\,dy, \tag{15a}$$

where $k(x, y)$ is a given symmetric function of x and y. For simplicity we shall assume that $\mathfrak{H}[\varphi]$ never takes negative values. In place of the eigenvalue differential equation we obtain, by the method of subsection 1, the *integro-differential equation*

$$(pu')' - qu + \lambda\left(\rho u + \int_G k(x, y)u(y)\,dy\right) = 0 \tag{16}$$

with the boundary condition $u = 0$, for example. The orthogonality relations[2] for the eigenfunctions of this problem become

$$\int_G \rho u_i(x)u_j(x)\,dx + \int_G\int_G k(x, y)u_i(x)u_j(y)\,dx\,dy = \begin{cases} 0 & \text{for} \quad i \neq j, \\ 1 & \text{for} \quad i = j. \end{cases}$$

[1] H. Kneser has suggested the term "weighted orthogonality."

[2] They can also be considered as "*biorthogonality relations*" between the system u_i and the system v_i if in addition to the eigenfunctions $u_i(x)$ we in-

3. Eigenvalue Problems for Regions with Separate Components. The following remark, valid for all eigenvalue problems of differential equations, will be of importance later:

If G consists of a number of non-overlapping (open) domains G', G'', $\cdots$, then the totality of eigenvalues and eigenfunctions of G consists of the eigenvalues and eigenfunctions of all the subdomains G', G'', $\cdots$, where each eigenfunction is understood to vanish identically in all except one of the separate subdomains.

This represents the self-evident physical fact that several separate vibrating systems perform vibrations without interacting with each other.

One way of proving the statement mathematically is to use the definition of the eigenfunctions by means of the differential equation problem. We need only remark that the eigenfunctions which are defined for any one of the subdomains G', G'', $\cdots$ and vanish identically outside this subdomain (together with linear combinations of such eigenfunctions belonging to the same eigenvalue) are also eigenfunctions of G. Conversely, every eigenfunction of G must differ from zero in at least one of the subregions. An alternative proof of the assertion is based on the definition of the eigenvalues by means of variational problems; proceeding step by step, it is easy to identify the eigenvalues of the total domain with those of the subdomains.

4. The Maximum-minimum Property of Eigenvalues. Again, as in our discussion of quadratic forms (cf. Chapter I), we can define the n-th eigenvalue and the associated eigenfunction without reference to preceding eigenvalues and eigenfunctions.

We consider any one of the preceding variational problems, retain the notation of subsection 1, and modify the problem by modifying the conditions for the admissible functions φ. Instead of stipulating $H[\varphi, u_i] = 0 \quad (i = 1, 2, \cdots, n-1)$ we impose the $n-1$ modified conditions

$$H[\varphi, v_i] = 0 \qquad (i = 1, 2, \cdots, n-1)$$

---

troduce the functions $v_i(x) = \rho u_i(x) + \int_a^b k(x, y)u_i(y)\,dy$. In other words, they can be written in the form

$$\int_a^b u_i v_j\,dx = \begin{cases} 0 & \text{for} \quad i \neq j, \\ 1 & \text{for} \quad i = j. \end{cases}$$

where $v_1, v_2, \cdots, v_{n-1}$ are arbitrarily chosen piecewise continuous functions in G. Under the conditions imposed, the integrals $D[\varphi]$, or more generally the functionals $\mathfrak{D}[\varphi]$, have a *greatest lower bound*, which depends on the functions $v_1, v_2, \cdots, v_{n-1}$ and will be denoted by $d\{v_1, v_2, \cdots, v_{n-1}\}$. We assert that the eigenfunctions u_n and the eigenvalues λ_n, which were defined successively by variational problems, can be characterized in terms of the modified problem by the following theorem:

Given $n-1$ functions $v_1, v_2, \cdots, v_{n-1}$ which are piecewise continuous in G, let $d\{v_1, v_2, \cdots, v_{n-1}\}$ be the greatest lower bound of the set of values assumed by the functional $\mathfrak{D}[\varphi]$, where φ is any function continuous in G with piecewise continuous derivatives which satisfies $H[\varphi] = 1$ and the conditions

$$H[\varphi, v_i] = 0 \qquad (i = 1, 2, \cdots, n-1) \tag{17}$$

Then λ_n is equal to the largest value which this lower bound d assumes if the functions $v_1, v_2, \cdots, v_{n-1}$ range over all sets of admissible functions. This maximum-minimum is attained for $u = u_n$ and $v_1 = u_1$, $v_2 = u_2, \cdots, v_{n-1} = u_{n-1}$.

Under the boundary condition $u = 0$, the variational problem is no longer a free problem but is restricted by the condition $\varphi = 0$ on Γ.

To prove the theorem, we first note that by definition we have $d\{v_1, v_2, \cdots, v_{n-1}\} = \lambda_n$ for $v_i = u_i$ $(1 \leq i \leq n-1)$; then, we show that for arbitrary $v_1, v_2, \cdots, v_{n-1}$ we have $d\{v_1, v_2, \cdots, v_{n-1}\} \leq \lambda_n$. We need only determine a particular function φ, satisfying the conditions $H[\varphi, v_i] = 0$ $(i = 1, 2, \cdots, n-1)$, for which $\mathfrak{D}[\varphi] \leq \lambda_n$. For this purpose we determine a suitable linear combination of the first n eigenfunctions $\varphi = \sum_{i=1}^{n} c_i u_i$ with constant $c_1, c_2, \cdots, c_n$. The $n-1$ relations (17) lead to $n-1$ linear homogeneous conditions on the n quantities $c_1, c_2, \cdots, c_n$, and thus can always be satisfied. The equation $H[\varphi] = \sum_{i=1}^{n} c_i^2 = 1$ is simply a normalizing condition which determines a factor of proportionality. Since $\mathfrak{D}[u_i, u_k] = 0$ $(i \neq k)$ and $\mathfrak{D}[u_i] = \lambda_i$ [cf. eq. (8)]

$$\mathfrak{D}[\varphi] = \sum_{i,k=1}^{n} c_i c_k \mathfrak{D}[u_i, u_k]$$

implies

$$\mathfrak{D}[\varphi] = \sum_{i=1}^{n} c_i^2 \lambda_i;$$

because of

$$\sum_{i=1}^{n} c_i^2 = 1 \quad \text{and} \quad \lambda_n \geq \lambda_i \qquad (i = 1, 2, \cdots, n),$$

we have

$$\mathfrak{D}[\varphi] \leq \lambda_n .$$

Thus the minimum $d\{v_1, v_2, \cdots, v_{n-1}\}$ is certainly no greater than λ_n; therefore, λ_n is the largest value which this minimum can assume.

§2. *General Consequences of the Extremum Properties of the Eigenvalues*

1. General Theorems. Important conclusions can be drawn from the maximum-minimum property by means of certain simple principles of the calculus of variations. The first of these principles states: *By strengthening the conditions in a minimum problem we do not diminish the value of the minimum*; *conversely, by weakening the conditions the minimum decreases, or at any rate does not increase.* The second principle states: *Given two minimum problems with the same class of admissible functions φ, such that for every φ the functional to be minimized is no smaller in the first problem than in the second, then the minimum for the first problem is also no smaller than the minimum for the second.*

The classical minimum definition of eigenvalues does not lend itself to the application of these principles when eigenvalues in different problems are compared; for, the classes of admissible functions fail to coincide since the conditions of admissibility are not the same. Under the maximum-minimum definition, however, the classes of admissible functions agree, and our principles may be applied.

Consider for example any system, capable of vibrating, the eigenvalues of which are determined by an eigenvalue problem of the type treated here. Then any conditions of constraint under which the system is required to vibrate may be expressed mathematically as auxiliary conditions imposed on the functions φ admissible in the variational problem. Whenever the conditions on φ in the maximum-minimum problem are strengthened, the lower bound $d\{v_1, v_2, \cdots, v_{n-1}\}$ increases, or at any rate does not decrease. Hence the same is true for the maximum of these lower bounds, the n-th

eigenvalue. Correspondingly, the value of the maximum-minimum, i.e. of the n-th eigenvalue, is diminished or at least not increased when the conditions on the functions φ are weakened.

Physically, this means:

THEOREM 1: *If a system which is capable of vibrating is subjected to constraining conditions, the fundamental tone and every overtone become higher (at least not lower) in pitch. Conversely, if restraining conditions are removed, the fundamental tone and every overtone become lower (at least not higher).*

For example, in the case of a vibrating stretched elastic membrane, the fundamental tone and all the overtones become higher if the membrane is fixed not only along the boundary but also along curves or at parts of the surface. On the other hand, the fundamental tone and all the overtones become lower if the membrane is slit or, in the case of a *vibrating plate*, if the material develops a crack. In this case the conditions of continuity for the admissible functions φ or their derivatives at the slit or crack are removed.

Mathematically, this principle leads to a number of important *general theorems on the distribution of eigenvalues.* Theorems 2 and 3 concern the boundary condition $u = 0$ and compare the eigenvalue distribution of a domain with that of subdomains. Further theorems make a corresponding statement for the boundary condition $\partial u/\partial n = 0$. Other theorems involve more general boundary conditions and compare the different spectra[1] of the differential equation for various forms of these boundary conditions.

THEOREM 2: *Suppose that G', G'', G''', $\cdots$ are a finite number of non-overlapping subdomains of the domain G. Let $A(\lambda)$ denote the number of eigenvalues less than λ of the differential equation $L[u] + \lambda\rho u = 0$ for G with the boundary condition $u = 0$. Then the total number of eigenvalues less than λ for all the separate subdomains with the same boundary condition does not exceed $A(\lambda)$.*

This theorem may also be expressed as follows: *Under the boundary condition $u = 0$, the n-th eigenvalue λ_n for the domain G is at most equal to the n-th number λ_n^* in the sequence consisting of all the eigenvalues of the subdomains $G^{(i)}$, arranged according to increasing magnitude and taken with their respective multiplicity.*

The proof follows immediately from the following consideration: Suppose that in the maximum-minimum problem which defines λ_n,

[1] By the term spectrum we mean, as before, the totality of eigenvalues.

we stipulate a new condition for the functions φ, requiring them to vanish at the boundaries of all the subdomains $G^{(i)}$ and to vanish in that part of G which belongs to no subdomain $G^{(i)}$. Then, first, in virtue of the fundamental principle stated above, the value of the maximum-minimum is not diminished. On the other hand, the resulting maximum-minimum is precisely that which defines the n-th eigenvalue of the disconnected region consisting of the separate domains G', G'', $\cdots$. In view of the remark in §1, 3, the new value of the maximum-minimum is equal to λ_n^*, and hence we have $\lambda_n \leq \lambda_n^*$, as we wished to prove.

In particular, the theorem just proved leads to an important property of *monotonicity* of the eigenvalues associated with the boundary condition $u = 0$.

THEOREM 3: *Under the boundary condition $u = 0$ the n-th eigenvalue for a domain G never exceeds the n-th eigenvalue for a subdomain of G.*[1]

A statement corresponding to that of Theorem 2 may be made for the boundary condition $\partial u/\partial n = 0$.

THEOREM 4: *Suppose that G', G'', G''', $\cdots$ are a finite number of non-overlapping subdomains which exhaust the domain G completely. Let $B(\kappa)$ denote the number of eigenvalues less than κ of the differential equation $L[u] + \lambda\rho u = 0$ for G with the boundary condition $\partial u/\partial n = 0$. Then the total number of eigenvalues less than κ for all the separate subdomains with the same boundary condition is at least as large as $B(\kappa)$.*

We can also state this theorem in the following way: *Let κ_n^* be the n-th number, arranged in order of increasing magnitude, in the combined set of eigenvalues belonging to the subdomains $G^{(i)}$, associated with the boundary condition $\partial u/\partial n = 0$, where every eigenvalue is to be taken with its correct multiplicity. Then the n-th eigenvalue κ_n of the domain G for the same boundary condition is larger than or equal to the number κ_n^*.*

Here, too, the proof follows almost immediately from the first of our general principles applied to the maximum-minimum problem characterizing the n-th eigenvalue κ_n for G. For, if we permit the admissible functions φ in this problem to be discontinuous on the boundary curves of the domains $G^{(i)}$ in G (the discontinuities across these boundary curves are at most finite jumps), we reduce the value of the maximum-minimum, or at least do not increase it. On the other hand, the modified maximum-minimum problem, according to

[1] In fact, it is always smaller when we are dealing with a proper subdomain, as can be ascertained easily by the method of §6.

§1, 4, defines just the n-th eigenvalue associated with the natural boundary conditions $\partial u/\partial n = 0$ for the domain consisting of the separate domains $G^{(i)}$, that is, the value κ_n^*. This proves the relation $\kappa_n \geq \kappa_n^*$.

We shall now prove some theorems concerning the relative behavior of the spectra of the differential equation for the different types of boundary conditions considered above.

THEOREM 5: *Let λ_n be the n-th eigenvalue of the differential equation $L[u] + \lambda\rho u = 0$ for the domain G under the boundary condition $u = 0$, and let μ_n be the n-th eigenvalue for the condition $\partial u/\partial n + \sigma u = 0$, or more generally for the condition $\partial u/\partial n + \sigma u = 0$ on a part Γ' of the boundary Γ, $u = 0$ on the remaining part Γ'' of the boundary. Then*

$$\mu_n \leq \lambda_n .$$

This is proved by considering the maximum-minimum problem which characterizes the n-th eigenvalue μ_n of G, without boundary conditions, as the maximum of the minimum of $\mathfrak{D}[\varphi]$; we impose on the functions φ the further condition of vanishing on the boundary Γ of G. Then the value of the individual minimum and hence also of the maximum-minimum is certainly increased or at least not diminished. On the other hand, this new maximum-minimum value is clearly identical with λ_n, since in view of the imposed condition we now have $\mathfrak{D}[\varphi] = D[\varphi]$. Hence $\mu_n \leq \lambda_n$, as was asserted.

THEOREM 6: *If, in the boundary condition $\partial u/\partial n + \sigma u = 0$ on Γ, the function σ is either increased or diminished at every point, then each individual eigenvalue can change only in the same sense.*

This important fact is likewise a direct consequence of the maximum-minimum property by the second of the above-mentioned principles. For, if we change the function σ, the expression $\mathfrak{D}[\varphi]$ changes in the same sense as σ for each φ; hence the same is true of its lower bound for given v_i, and therefore for the maximum of these lower bounds.

Theorems 5 and 6 exhibit a mutual relationship between the eigenvalues corresponding to the various types of boundary conditions. If we change the function σ at each point monotonically from 0 to ∞, each individual eigenvalue μ increases monotonically from the value which it had for the boundary condition $\partial u/\partial n = 0$ to the value which it takes for the boundary condition $u = 0$. In other words, the theorem states that *of the boundary conditions considered*, $u = 0$

is the most restrictive and $\partial u/\partial n = 0$ *is the least restrictive if* σ *is non-negative.* That the limit of the eigenvalue μ_n for infinitely increasing σ is indeed λ_n may best be proved by investigating the nature of the eigenfunctions more closely. Since this will not be done until later, we refrain from carrying out the proof at this point (cf. Volume II).

In subsection 6 we shall see that this growth with increasing σ is continuous. Investigation of the asymptotic distribution of the eigenvalues for large n will show, furthermore, that in spite of the indicated behavior of the eigenvalues, their asymptotic behavior for $n \to \infty$ is independent of the boundary condition, and that accordingly the growth of the n-th eigenvalue due to the growth of the function σ is negligibly small compared with the size of the eigenvalue for sufficiently large n.

The properties formulated in Theorems 5 and 6 suggest a simple *physical interpretation.* The boundary condition $\partial u/\partial n + \sigma u = 0$ describes a boundary fixed by elastic forces, where the magnitude of the constraining forces is determined by the function σ. The theorems state that as this elastic constraint increases in intensity each eigenfrequency increases. The condition $u = 0$ represents the case in which this force has become infinite, or in other words the boundary is completely fixed.

Finally, the maximum-minimum property of the eigenvalues makes it possible to investigate the dependence of the eigenvalues on the coefficients of the differential equation and on the domain G.

THEOREM 7: *If, in the differential equation* $L[u] + \lambda\rho u = 0$, *the coefficient* ρ *varies at every point in the same sense, then, for every boundary condition, the n-th eigenvalue changes in the opposite sense. If either of the coefficients* p *or* q *changes everywhere in the same sense, every eigenvalue changes in this same sense.* (In the case of the boundary condition $\partial u/\partial n + \sigma u = 0$, we assume $\sigma \geq 0$.)

First let p be changed everywhere in the same sense. Then for every admissible function φ the value of the expression $\mathfrak{D}[\varphi]$, therefore also the lower bounds of these values for fixed v_i and the maximum of these lower bounds, the n-th eigenvalue, change monotonically in the same sense. If ρ is changed monotonically, resulting in the new function $\rho' \geq \rho$, then for any admissible function φ we have

$$\mathfrak{D}[\varphi]: \iint_G \rho\varphi^2\, dx\, dy \geq \mathfrak{D}[\varphi]: \iint_G \rho'\varphi^2\, dx\, dy.$$

Here we see that for fixed functions v_i the lower bound of the left side is no smaller than the lower bound of the right side. In forming the lower bound of the latter quotient we have to substitute the functions $v_i' = v_i\rho/\rho'$ for the functions v_i in the auxiliary conditions, since ρ has been changed to ρ'. If the system of all the functions v_i ranges over the entire class of admissible functions then the system of functions v_i' does the same; it is, therefore, clear that the ratio of the sizes of the maxima of the lower bounds considered is opposed to that of the functions ρ and ρ'.

2. Infinite Growth of the Eigenvalues. In this subsection we shall show that in the *variational eigenvalue problems the eigenvalues* λ_n *become infinite for* $n \to \infty$. This implies, in particular, that each eigenvalue has only a finite multiplicity, and that only a finite number of eigenvalues can be negative. The most important consequence of the unboundedness of the eigenvalues is, as will be seen in §3, 1, the completeness of the system of eigenfunctions; this system coincides therefore with the system of eigenfunctions of the differential equation.

To prove the unboundedness (here we shall not assume $q > 0$) we denote by p_M, q_M, ρ_M and p_m, q_m, ρ_m the largest and smallest values respectively of the functions p, q, ρ in G, and consider first the boundary condition $u = 0$. If we replace the functions p, q, ρ in $\mathfrak{D}$ and H by the constants p_m, q_m, ρ_M and p_M, q_M, ρ_m, we obtain new eigenvalue problems with the eigenvalues λ_n' and λ_n'' respectively, and by Theorem 7 we have $\lambda_n' \leq \lambda_n \leq \lambda_n''$.

We ascertain first that the eigenvalues λ_n' are unbounded: In the case of one independent variable, we can solve the associated differential equation eigenvalue problem explicitly in terms of trigonometric functions; the eigenvalues are the numbers $(p_m\nu^2 + q_m)/\rho_M$, $\nu = 1, 2, \cdots$. Since the eigenvalues λ_n' arising from the variational problem are certainly contained in this sequence, our assertion $\lambda_n \to \infty$ follows immediately.

We stated above that the eigenvalues resulting from the variational problem coincide with the totality of the eigenvalues of the differential equation; assuming for the present that this assertion is true, we have

$$\lambda_n' = \frac{p_m n^2 + q_m}{\rho_M}, \qquad \lambda_n'' = \frac{p_M n^2 + q_M}{\rho_m},$$

and thus we see that the quotient λ_n/n^2 remains between finite positive bounds as n increases.

In order to estimate the eigenvalues λ'_n for an arbitrary domain G in several dimensions, we compare them with the eigenvalues λ^*_n of the same differential equation for a square containing all of G. We know from Ch. V, §14 that these become infinite as n increases; the same is true for λ_n, since by Theorems 3 and 7 we have

$$\lambda^*_n \leq \lambda'_n \leq \lambda_n .$$

We shall not extend this reasoning to different boundary conditions since more precise estimates of the asymptotic behavior of the eigenvalues will soon show that they are unbounded. However, we sketch briefly an entirely different—indirect—method of proof which presupposes no knowledge of the solutions of the variational problem in special cases, and thus is intrinsically preferable.[1]

In the case of one independent variable, we assume that our problem has infinitely many eigenvalues $\lambda_1, \lambda_2, \cdots$ whose absolute values all remain below some positive bound. The boundedness of

$$\lambda_n = \int_{x_1}^{x_2} (pu_n'^2 + qu_n^2)\, dx + h_1 p(x_1)u_n(x_1)^2 + h_2 p(x_2)u_n(x_2)^2$$

and of $\int_{x_1}^{x_2} \rho u_n^2\, dx$ immediately implies the boundedness of

$$\int_{x_1}^{x_2} u_n'^2\, dx \quad \text{and of} \quad \int_{x_1}^{x_2} u_n^2\, dx,$$

if the constants h_1 and h_2 are non-negative. (This restriction can be easily removed with the help of the observations in subsection 5.)

We now make use of the following lemma: If for a set of functions $\varphi(x)$ the integrals $\int_G \varphi'^2\, dx$ and $\int_G \varphi^2\, dx$ are bounded, the functions $\varphi(x)$ are equicontinuous and uniformly bounded (cf. Ch. II, p. 58.) Thus by the accumulation principle (Ch. II, §2) we can select a uniformly convergent subsequence from the set of eigenfunctions u_n. If we again denote this new sequence by u_n, it is clear that $\lim_{n,m\to\infty} H[u_n - u_m] = 0$. But on the other hand, from the orthogonality of the u_n for $n \neq m$, we have

$$H[u_n - u_m] = 2.$$

This contradiction proves the theorem.

In the case of several (e.g. two) variables, the same reasoning holds

[1] This method is due to F. Rellich, Ein Satz über mittlere Konvergenz, Nachr. Ges. Göttingen (math.-phys. Kl.), 1930.

on the basis of the following lemma by Rellich[1] (which will not be proved here):

Given a set of functions $\varphi(x, y)$ defined in the domain G, we can select a subsequence φ_n for which

$$\lim_{n,m\to\infty} \iint_G (\varphi_n - \varphi_m)^2 \, dx \, dy = 0$$

if the expressions $\iint_G (\varphi_x^2 + \varphi_y^2) dx \, dy$ *and* $\iint_G \varphi^2 \, dx \, dy$ *are uniformly bounded.*

3. Asymptotic Behavior of the Eigenvalues in the Sturm-Liouville Problem. In the case of the Sturm-Liouville problem, the maximum-minimum property makes it possible both *to determine the order of magnitude of the n-th eigenvalue and to find its asymptotic value.* Using the transformation (20a) of Ch. V, §3, 3, we write the differential equation $(py')' - qy + \lambda \rho y = 0$ in the form

$$z'' - rz + \lambda z = 0 \tag{18}$$

for the interval $0 \leq t \leq l$, where $r(t)$ is a continuous function and z must satisfy new homogeneous boundary conditions which arise from the original ones. We consider first the case $y(0) = y(\pi) = 0$ or $z(0) = z(l) = 0$. The eigenvalues are maxima of the minima of an expression of the form

$$\int_0^l (z'^2 + rz^2) \, dx$$

—in this subsection we shall assume the fact, to be proved later (§3, 1), that the eigenfunctions and eigenvalues arising from the variational problem coincide with those associated with the differential equation. If we omit the term rz^2 in this expression, and thus consider the integral expression

$$\int_0^l z'^2 \, dx,$$

then, provided z satisfies the condition

$$\int_0^l z^2 \, dx = 1,$$

[1] See Rellich, op. cit.

the new expression differs from the original one by no more than a fixed finite bound r_M (maximum of the absolute value of r). Hence the maximinima of the first expression, i.e. the desired eigenvalues, differ from those of the second by no more than r_M. But the maximinima of the second integral are just the eigenvalues $\mu_n = n^2\pi^2/l^2$ of $z'' + \mu z = 0$ for the interval $(0, l)$; therefore, since $\lim_{n\to\infty} \mu_n = \infty$, we immediately have *the asymptotic formula*

$$\lambda_n = \mu_n + O(1), \tag{19}$$

where, as before, $O(1)$ denotes a function which remains bounded as n becomes infinite. Returning to our original notation, we find

$$\lim_{n\to\infty} \frac{n^2}{\lambda_n} = \frac{1}{\pi^2}\left(\int_0^\pi \sqrt{\frac{\rho}{p}}\, dx\right)^2. \tag{19a}$$

Precisely the same estimate may be derived for arbitrary boundary conditions, since the asymptotic behavior of the eigenvalues of the equation $z'' + \mu z = 0$ is independent of the boundary condition (cf. also §4.)

4. Singular Differential Equations. Our asymptotic estimate is easily carried over to differential equations with singularities. We consider only Bessel's equation

$$xu'' + u' + \left(x\lambda - \frac{m^2}{x}\right)u = 0,$$

whose solutions are the Bessel functions $J_m(x\sqrt{\lambda})$. Here, with the boundary conditions $u(0) < \infty$ and $u(1) = 0$, the eigenvalues λ are the squares $\lambda_{m,n}$ of the zeros of J_m (cf. Ch. V, §10, 1). In the case $m \geq 1$ it is useful to consider the functions $v = \sqrt{x}J_m(x\sqrt{\lambda})$, with the eigenvalue equation

$$v'' + \left(\lambda - \frac{4m^2 - 1}{4x^2}\right)v = 0$$

and to characterize its eigenvalues by maxima of minima of the quotient $D[\varphi]/H[\varphi]$ with

$$D[\varphi] = \int_0^1 \left(\varphi'^2 + \frac{4m^2 - 1}{4x^2}\,\varphi^2\right) dx, \qquad H[\varphi] = \int_0^1 \varphi^2\, dx;$$

this replaces the characterization given in §1, 2. As a boundary condition we set $\varphi(0) = \varphi(1) = 0$; since $m \geq 1$, we certainly have

$D[\varphi] \geq \int_0^1 \varphi'^2\, dx$ and thus $\lambda_n \geq n^2\pi^2$. On the other hand we obtain an upper bound for λ_n by restricting the admissibility conditions in the following way: In an interval $0 \leq x \leq \epsilon$ to be determined later, we impose the condition $\varphi(x) = 0$ and majorize the second term in the integrand of $D[\varphi]$ by the constant

$$\frac{4m^2 - 1}{4\epsilon^2} \int_0^1 \varphi^2\, dx = \frac{c}{\epsilon^2}.$$

This yields $\lambda_n \leq n^2\pi^2/(1 - \epsilon)^2 + c/\epsilon^2$. Now if we let ϵ approach zero with $1/n$ in a suitable way, e.g., if we set $\epsilon = 1/\sqrt{n}$, it follows that $\lim_{n\to\infty} \lambda_n/n^2\pi^2 \leq 1$. We thus obtain the asymptotic formula

$$\lim_{n\to\infty} \frac{\lambda_n}{n^2\pi^2} = 1$$

for the zeros $\sqrt{\lambda_{m,n}}$ of J_m, which agrees with the formulas for the nonsingular problems. The same asymptotic formula holds for other conditions which we have considered, e.g. for the boundary condition $u'(1) = 0$.

This result may be immediately extended to the roots of the Bessel function of zero-th order, in view of the relation $J_0'(x) = -J_1(x)$ (cf. Ch. V, §5, 5). Accordingly, the eigenvalues of Bessel's problem with the boundary condition $u(1) = 0$ for $m = 1$ are identical with those for the boundary condition $u'(1) = 0$ for $m = 0$ (apart from the first eigenvalue which is zero). Since $J_0(x)$ has exactly one zero between two consecutive zeros of $J_0'(x)$—as we may prove with the aid of Rolle's theorem in the manner of Ch. VII, §2, 8—this justifies the asymptotic formula for $m = 0$.[1]

5. Further Remarks Concerning the Growth of Eigenvalues. (Occurrence of) Negative Eigenvalues. If, as we have assumed, the function σ and the numbers h_1, h_2 in the variational problems of §1 are non-negative,[2] and the same is true of q, then clearly there can be no negative eigenvalues. Now the considerations of subsection 2 show: *If the function q is not everywhere positive, at most a finite number of negative eigenvalues can occur. This is also true when the function σ or the constants h_1, h_2 assume negative values* since here, too, the eigenvalues become infinite with increasing n.

[1] See another proof for $m = 0$ in §7, 8.

[2] The function p, as well as ρ, is always assumed non-negative.

To prove this, we consider, for the sake of brevity, a one-dimensional problem with the fundamental domain $0 \leq x \leq \pi$ and estimate the negative terms associated with the boundary: If ξ denotes a point in the interval $0 \leq x \leq t$, we have

$$|y(0) - y(\xi)| = \left| \int_0^\xi y' \, dx \right|,$$

where t is to be chosen subject only to the restriction $0 < t \leq \pi$. Using the Schwarz inequality we find

$$|y(0) - y(\xi)| \leq \sqrt{t} \sqrt{\int_0^\pi y'^2 \, dx}$$

and hence, denoting the minimum of p by p_m,

$$|y(0)| \leq |y(\xi)| + \sqrt{\frac{t}{p_m}} \sqrt{\int_0^\pi py'^2 \, dx}.$$

Under the condition $\int_0^\pi \rho y^2 \, dx = 1$ there exists an intermediate point ξ such that $y(\xi)^2 \leq 1/t\rho_m$, where ρ_m denotes the minimum of ρ. Thus we have

$$|y(0)| \leq \frac{1}{\sqrt{t\rho_m}} + \sqrt{\frac{t}{p_m}} \sqrt{\int_0^\pi py'^2 \, dx}.$$

If the integral under the radical sign exceeds the fixed bound $1/\pi^2$, we require t to satisfy

$$\frac{1}{t} = \sqrt{\int_0^\pi py'^2 \, dx}.$$

In this case t falls in the interval $0 \leq x \leq \pi$; otherwise we set $t = \pi$. Then we have

$$y(0)^2 \leq c \sqrt{\int_0^\pi py'^2 \, dx} + c_1,$$

where c and c_1 are constants independent of $y(x)$. It follows immediately that since a similar estimate is valid for $y(\pi)$, every admissible function y satisfies the important relation

$$h_1 y(0)^2 + h_2 y(\pi)^2 \leq C_1 \sqrt{\int_0^\pi py'^2 \, dx} + C_2,$$

where C_1, C_2 are suitable constants. Furthermore, we certainly have

$$\left| \int_0^\pi qy^2 \, dx \right| \leq C_3$$

and thus finally

$$\mathfrak{D}[y] \geq \int_0^\pi py'^2 \, dx - C_4 \sqrt{\int_0^\pi py'^2 \, dx} - C_5 \geq \frac{1}{2} \int_0^\pi py'^2 \, dx - C_6.$$

Since the eigenvalues associated with the integral $\int_0^\pi py'^2 \, dx$ become infinite, those belonging to $\mathfrak{D}[y]$ also become infinite. Therefore there can be only a finite number of negative eigenvalues.

In an analogous way we obtain for two-dimensional problems an estimate of the form

$$\left| \int_\Gamma p\sigma\varphi^2 \, ds \right| \leq c_1 \sqrt{|D[\varphi]|} + c_2, \tag{20}$$

from which the same results concerning the essentially positive character of the eigenvalues follow.[1] We mention finally that entirely analogous considerations show the unboundedness of the eigenvalues for the general eigenvalue problems of §1, 2.[2]

6. Continuity of Eigenvalues. Assume first that the function ρ is changed to a function ρ', so that $0 < (1 - \epsilon)\rho \leq \rho' \leq (1 + \epsilon)\rho$ for positive ϵ. Then by Theorem 7, the n-th eigenvalue of the differential equation lies between the n-th eigenvalues obtained by replacing ρ by $\rho(1 - \epsilon)$ and by $\rho(1 + \epsilon)$, respectively, in the differential equation, i.e. by multiplying the n-th eigenvalue of the original differential equation by the factors $(1 - \epsilon)^{-1}$ and $(1 + \epsilon)^{-1}$, respectively. If ϵ is taken sufficiently small, then these two numbers lie arbitrarily close to each other. *Thus the n-th eigenvalue depends continuously on the function* ρ.

The n-th eigenvalue also depends continuously on q. In fact, from the relation $\rho \geq \rho_m$, where ρ_m is a positive constant, it follows that

$$1 = \iint_G \rho\varphi^2 \, dx \, dy \geq \rho_m \iint_G \varphi^2 \, dx \, dy.$$

[1] See Courant, Über die Eigenwerte bei den Differentialgleichungen der mathematischen Physik, in particular pp. 13–17.

[2] See Courant, Über die Anwendung der Variationsrechnung. . . .

Clearly, the integral $\iint_G \varphi^2\,dx\,dy$ is uniformly bounded for all admissible functions φ. This implies that for a sufficiently small change in the function q the expression $\mathfrak{D}[\varphi]$ changes arbitrarily little; in fact, it changes uniformly for all admissible functions φ. The same is therefore true for the maximum-minimum of $\mathfrak{D}[\varphi]$.

In a similar way the eigenvalues are seen to depend continuously on the function σ which occurs in the boundary condition. We may again suppose that the expressions $\mathfrak{T}[\varphi]$ in the variational problem have a fixed upper bound.[1] Then by relation (20) the boundary integrals $\int_\Gamma p\sigma\varphi^2\,ds$ have a common bound. Thus if we change the function σ in the boundary integral $\int_\Gamma p\sigma\varphi^2\,ds$ by a sufficiently small amount, then the change in the integral is also arbitrarily small and uniformly small for all admissible φ. Therefore the same is true for $\mathfrak{D}[\varphi]$ and for the maximum-minimum of $\mathfrak{D}[\varphi]$.

Continuous dependence on p follows in a similar manner.

Summing up, we have:

THEOREM 8: *For all boundary conditions considered, the n-th eigenvalue of the differential equation* $L[u] + \lambda\rho u = 0$ *depends continuously on the coefficients of the equation.*

THEOREM 9: *The n-th eigenvalue depends continuously on the function* σ *which occurs in the boundary condition* $\partial u/\partial n + \sigma u = 0$.

Finally, we investigate the continuity properties of *the n-th eigenvalue, considered as a function of the domain G.* It will be seen that the n-th eigenvalue of a domain G', for corresponding boundary conditions, approximates the n-th eigenvalue of the domain G arbitrarily closely, if the domain G' approximates G sufficiently closely. *We must, however, use a sufficiently strong concept of approximation of a domain G by another domain G'.* Whenever normal derivatives occur in the boundary conditions, it will not be enough to require that the boundary of G' approximate the boundary of G pointwise. We shall

[1] For instance, the upper bound may be the n-th eigenvalue for an arbitrary square lying entirely in the interior of G with the boundary condition $u = 0$. For, Theorems 3 and 5 imply that the n-th eigenvalue for G with the original boundary condition is certainly no greater than the n-th eigenvalue for such a square with the boundary condition $u = 0$. Thus we see that fixing this upper bound for $\mathfrak{D}[\varphi]$ cannot affect the solution of the maximum-minimum problem.

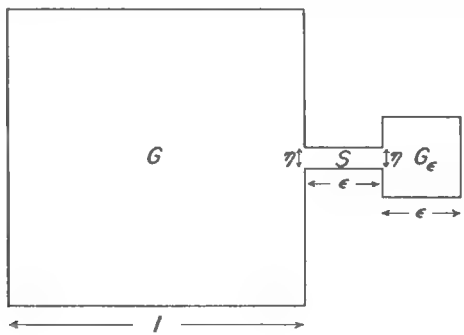
have to require, in addition, that the *normals* to the boundary of G' *approximate* those of G. It can be shown, in fact, that under the weaker definition of approximation the n-th eigenvalue need not be a continuous function of the domain.[1]

Approximation of the domain G by the domain G' in the stronger sense is defined analytically in the following way:

Let G together with its boundary be transformed pointwise into the domain G' together with its boundary by equations of the form

$$x' = x + g(x, y), \qquad y' = y + h(x, y); \tag{21}$$

[1] The following is a simple counterexample: Let $L[\varphi] = \Delta\varphi$, $\rho = 1$, and let G be a square of edge 1. Outside G we take a second square G_ϵ of side ϵ, oriented parallel to G, which lies opposite one of the sides of G and at a distance ε from it. We then connect its interior with that of G by means of a narrow strip S, perpendicular to the two squares, formed by two parallel line-segments of length ϵ at a distance η from each other. Let the domain G' consist of the two squares G and G_ϵ together with the strip S. The first eigenvalue of G' with the boundary condition $\partial u/\partial n = 0$ is zero, the associated eigenfunction is $u_1 = \text{const}$. Now if, for every ϵ, the width η of the strip S is taken sufficiently small, then the second eigenvalue of G' can also be made



$$G' = G + G_\epsilon + S$$

arbitrarily small. For, consider a function φ in G' which is equal to $-1/\epsilon$ in G_ϵ and to a constant c in G, and which falls off linearly in S from c to $-1/\epsilon$. Let the constant c be chosen in such a way that the integral of φ over G' vanishes. If ϵ is sufficiently small, c is arbitrarily close to 0. The integral $\mathfrak{D}[\varphi]$ over G' will be of the order η/ϵ^3. If, for example, we choose $\eta = \epsilon^4$, this integral is arbitrarily small while the integral of φ^2 over G' is arbitrarily close to 1. Hence, as a consequence of the classical minimal property of the eigenvalues and eigenfunctions, the second eigenvalue of G' is certainly arbitrarily small. If we permit ϵ to approach zero, the second eigenvalue of G' will certainly converge to zero if η/ϵ^3 does. But the second eigenvalue of G is positive; hence it is not the limit of the second eigenvalue of G' although the boundary of G' converges to the boundary of G.

here the functions g, h are continuous and have piecewise continuous first derivatives throughout the domain, and both g, h and their first derivatives are less in absolute value than a small positive number ϵ. Then we say *that the domain G is approximated by the domain G' with the degree of accuracy ϵ.*

When ϵ approaches zero we say that G' is deformed continuously into G. We shall prove the following theorem:

THEOREM 10: *For any of the boundary conditions considered, the n-th eigenvalue of the differential equation $L[u] + \lambda\rho u = 0$ varies continuously when the domain G is deformed continuously in the sense defined above.*

Proof: Consider a sequence of domains G' for which the above numbers ϵ converge to zero. We solve equations (21) for x and y, set

$$\varphi(x, y) = \varphi'(x', y'), \quad p(x, y) = p'(x', y') \text{ etc.}, \quad \sigma(x(s), y(s)) = \tau(s),$$

and transform the two integrals which constitute $\mathfrak{D}[\varphi]$ into an integral over the domain G' and one over its boundary Γ'.

In this way we obtain an eigenvalue variational problem for the domain G, with coefficients differing arbitrarily little from the original coefficients. Accordingly, continuous dependence may be proved by methods analogous to those which led to Theorems 8 and 9. We shall carry out the proof in detail:

The integral $D[\varphi]$ is transformed into the integral

$$\iint_{G'} \{p'[(\varphi'_{x'}(1 + g_x) + \varphi'_{y'}h_x)^2 + (\varphi'_{x'}g_y + \varphi'_{y'}(1 + h_y))^2] + q'\varphi'^2\} M^{-1}dx'dy', \tag{22}$$

where M stands for the functional determinant

$$M = \left(1 + \frac{\partial g}{\partial x}\right)\left(1 + \frac{\partial h}{\partial y}\right) - \frac{\partial g}{\partial y}\frac{\partial h}{\partial x}$$

which is arbitrarily close to 1 if ϵ is sufficiently small. The integral over the boundary becomes

$$\int_{\Gamma} p\sigma\varphi^2\, ds = \int_{\Gamma'} p'\tau(s)\varphi'^2 \frac{ds}{ds'}\, ds',$$

where ds' denotes the element of arc on the boundary Γ' of G'.

If in general we define

$$D'[\psi] = \iint_{G'} \{p(\psi_x^2 + \psi_y^2) + q\psi^2\}\, dx\, dy, \quad \mathfrak{D}'[\psi] = D'[\psi] + \int_{\Gamma'} p\tau(s)\psi^2\, ds',$$

then the integrand in (22) differs from the integrand of $D'[\varphi']$ by the factors M^{-1} and p/p' (which differ arbitrarily little from 1) and by the additive terms in which $\varphi_{x'}'^2$, $\varphi_{y'}'^2$, $\varphi_{x'}'$, $\varphi_{y'}'$, and φ'^2, multiplied by factors which converge to zero with ϵ, appear. Using the inequality

$$2\left|\iint_{G'} \varphi_{x'}'\,\varphi_{y'}'\,dx'\,dy'\right| \leq \iint_{G'} (\varphi_{x'}'^2 + \varphi_{y'}'^2)\,dx'\,dy',$$

we obtain the relation

$$D[\varphi] = (1 + \delta)\,D'[\varphi'],$$

where δ denotes a quantity which approaches zero with ϵ. This notation will be used in the sequel, although δ will not always stand for the same quantity. Since ds/ds' is arbitrarily close to 1, for sufficiently small ϵ, we have

$$\int_{\Gamma'} p'\tau(s)\varphi'^2\,\frac{ds}{ds'}\,ds' = (1 + \delta)\int_{\Gamma'} p\tau(s)\varphi'^2\,ds'$$

and therefore

$$\mathfrak{D}[\varphi] = (1 + \delta)\,\mathfrak{D}'[\varphi'].$$

Furthermore, we must transform the auxiliary conditions (3a) and (17) imposed on the functions φ in §1. Thus we obtain the expressions

$$\iint_G \rho\varphi^2\,dx\,dy = \iint_{G'} \rho' M^{-1}\varphi'^2\,dx'\,dy' = 1,$$

$$\iint_G \rho\varphi v_i\,dx\,dy = \iint_{G'} \rho' M^{-1}\varphi' v_i'\,dx'\,dy' = 0 \qquad (i = 1, 2, \cdots, n-1).$$

We multiply the function φ' by a constant factor and the functions v_i' by $M^{-1}\rho'/\rho$; these factors all differ arbitrarily little from 1 for small ϵ. The resulting functions φ'' and v_i'' satisfy the relations

$$\iint_{G'} \rho\varphi''^2\,dx'\,dy' = 1,$$

$$\iint_{G'} \rho\varphi'' v_i''\,dx'\,dy' = 0 \qquad (i = 1, 2, \cdots, n-1).$$

We find that

$$\mathfrak{D}[\varphi] = (1 + \delta)\,\mathfrak{D}'[\varphi'']$$

holds and that the function φ'' satisfies the conditions of the maximum-minimum problem which characterizes the n-th eigenvalue for G'; here the functions v_i'' in G' correspond to the functions v_i in G. The domain of variability of the functions v_i'' is just the set of admitted functions over G'; it follows that the maximum-minimum of the left side can differ from that of the right only by a factor which approaches 1 as ϵ approaches zero, which proves Theorem 10. The above reasoning results in the following sharper form of this theorem:

COROLLARY TO THEOREM 10: *If a domain G is deformed into a domain G' by the transformation* (21) *in such a way that*

$$\left|\frac{\partial g}{\partial x}\right| < \epsilon, \qquad \left|\frac{\partial g}{\partial y}\right| < \epsilon, \qquad \left|\frac{\partial h}{\partial x}\right| < \epsilon, \qquad \left|\frac{\partial h}{\partial y}\right| < \epsilon,$$

where ϵ is any small positive number, then there exists a number η, depending only on ϵ and approaching zero with ϵ, such that for every n and for any one of the boundary conditions in question the n-th eigenvalues μ_n, μ_n' for the domains G and G', respectively, satisfy the relation

$$\left|\frac{\mu_n'}{\mu_n} - 1\right| < \eta.$$

For the boundary condition $u = 0$, in which no normal derivatives appear, the theorem on continuous dependence holds under less restrictive conditions:

THEOREM 11: *For the boundary condition $u = 0$, the n-th eigenvalue of the equation $L[u] + \lambda \rho u = 0$ is a continuous function of the domain G, even though the normals are no longer required to vary continuously when the domain is continuously deformed.*

In fact, we can always enclose the boundaries of any two domains G and G', which are sufficiently close together (but for which the normals at neighboring boundary points do not necessarily have neighboring directions), between the boundaries of two domains B and B' which are sufficiently close in the more restrictive sense. Since by Theorem 3 the n-th eigenvalue for the boundary condition $u = 0$ is a monotonic function of the domain, the n-th eigenvalues of G and G' must lie between those of B and B', which by Theorem 10 are themselves close together. This proves Theorem 11.

If we do not pass to the limit $\epsilon \to 0$, the preceding considerations furnish the following more general result:

If two domains G and G' are related to each other by a point trans-

formation of the above type such that the absolute value of the functional determinant is uniformly bounded, and if λ_n *and* λ_n' *denote the n-th eigenvalues of the domains G and G', respectively, then for sufficiently large n the quotient* λ_n/λ_n' *lies between positive bounds which are independent of n.*

§3. *Completeness and Expansion Theorems*

1. Completeness of the Eigenfunctions. In connection with the variational problems for the quotients $\mathfrak{D}[\varphi]/\mathfrak{H}[\varphi]$ considered in §§1 and 2, we obtained the relation

$$\lim_{n\to\infty} \lambda_n = \infty.$$

For the result it was essential that $\mathfrak{H}[\varphi]$ be positive definite, i.e. that $\mathfrak{H}[\varphi]$ can assume only non-negative values and vanishes only if the integrand φ vanishes. Using the above limit relation we shall now prove the completeness theorem in a sharper form.

The system of eigenfunctions associated with the quotient $\mathfrak{D}[\varphi]/\mathfrak{H}[\varphi]$ *is complete in the following sense: For any continuous function f and any positive* ϵ*, arbitrarily small, we can find a finite linear combination*

$$\alpha_1 u_1 + \alpha_2 u_2 + \cdots + \alpha_n u_n = \omega_n$$

of the eigenfunctions such that

$$\mathfrak{H}[f - \omega_n] < \epsilon.$$

We obtain the best approximation, i.e. the smallest value of $\mathfrak{H}[f - \omega_n]$*, with the Fourier coefficients*

$$\alpha_i = c_i = \mathfrak{H}[f, u_i].$$

These satisfy the completeness relation

$$\mathfrak{H}[f] = \sum_{i=1}^{\infty} c_i^2. \tag{23}$$

Here, as in the case of arbitrary orthogonal functions, it follows from relations (8) of §1 that the best mean approximation of f with respect to $\mathfrak{H}$ by a linear combination of the first n eigenfunctions, that is, the smallest value of $\mathfrak{H}[f - \omega_n]$, is attained when $\alpha_i = c_i = \mathfrak{H}[f, u_i]$ (the coefficients are independent of n). The relation

$$0 \leq \mathfrak{H}\left[f - \sum_{i=1}^{n} c_i u_i\right] = \mathfrak{H}[f] - \sum_{i=1}^{n} c_i^2$$

immediately implies the convergence of the infinite series $\sum_{i=1}^{\infty} c_i^2$ or, more precisely, the Bessel inequality $\sum_{i=1}^{\infty} c_i^2 \leq \mathfrak{H}[f]$.

To prove not only this inequality but the completeness relation (23), we assume first that the function f satisfies the admissibility conditions of the variational problem. Then the function

$$\rho_n = f - \sum_{i=1}^{n} c_i u_i$$

satisfies the orthogonality relations

$$\mathfrak{H}[\rho_n, u_i] = 0 \qquad (i = 1, 2, \cdots, n) \tag{24}$$

and, by §1 equation (7), also the orthogonality relations

$$\mathfrak{D}[\rho_n, u_i] = 0 \qquad (i = 1, 2, \cdots, n). \tag{24'}$$

From (24) and from the minimal property of λ_{n+1}, we find

$$\lambda_{n+1}\mathfrak{H}[\rho_n] \leq \mathfrak{D}[\rho_n]. \tag{25}$$

On the other hand, $\mathfrak{D}[\rho_n]$ is bounded, since

$$\mathfrak{D}[f] = \mathfrak{D}\left[\sum_{i=1}^{n} c_i u_i\right] + 2\mathfrak{D}\left[\sum_{i=1}^{n} c_i u_i, \rho_n\right] + \mathfrak{D}[\rho_n];$$

therefore, from relations (24') we have

$$\mathfrak{D}[f] = \mathfrak{D}\left[\sum_{i=1}^{n} c_i u_i\right] + \mathfrak{D}[\rho_n].$$

$\mathfrak{D}[\sum_{i=1}^{n} c_i u_i]$ equals $\sum_{i=1}^{n} \lambda_i c_i^2$ and remains greater than a fixed bound as n increases since only a finite number of eigenvalues can be negative. Thus $\mathfrak{D}[\rho_n]$ is bounded from above.

In virtue of relation (25) and the infinite growth of λ_{n+1},

$$\mathfrak{H}[f] - \sum_{i=1}^{n} c_i^2 = \mathfrak{H}[\rho_n] \to 0 \quad \text{for} \quad n \to \infty,$$

which proves the completeness relation (23) and therefore the completeness of the eigenfunctions.

If the continuous function f fails to satisfy the admissibility conditions of the problem, we can approximate it by a function f^* which does satisfy these conditions in such a way that $\mathfrak{H}[f - f^*] < \epsilon/4$, and then approximate f^* by a function $f_n^* = \sum_{i=1}^{n} c_i^* u_i$ with the property $\mathfrak{H}[f^* - f_n^*] < \epsilon/4$. Then, since

$$\mathfrak{H}[f - f_n^*] = \mathfrak{H}[f - f^*] + \mathfrak{H}[f^* - f_n^*] + 2\mathfrak{H}[f - f^*, f^* - f_n^*],$$

we have from the Schwarz inequality $\mathfrak{H}[f - f_n^*] < \epsilon$ and from the minimal property of ρ_n, $\mathfrak{H}[\rho_n] < \epsilon$. This proves the completeness theorem for functions which are only continuous.

The completeness of the system of solutions of the variational problem implies that these solutions actually represent the totality of eigenfunctions of the corresponding differential equation problem (cf. the reasoning frequently used in Chapter V, e.g. page 301).

From the completeness relation (23) we easily deduce the more general completeness relation for a pair of functions f and g,

$$\mathfrak{H}[f, g] = \sum_{i=1}^{\infty} \mathfrak{H}[f, u_i]\mathfrak{H}[g, u_i]. \tag{23a}$$

By a slight addition to our argument, relation (23) can be supplemented by

$$\mathfrak{D}[f] = \sum_{i=1}^{\infty} \lambda_i c_i^2 \tag{23b}$$

provided $\mathfrak{D}[f]$ is finite.

We assume first that a continuous $L[f] = g$ exists, with $\mathfrak{H}[g]$ finite. Then using Green's formula, we write

$$\mathfrak{D}[f] = -\mathfrak{H}[f, g].$$

Applying the completeness relation (23a) to the inner product $\mathfrak{H}[f, g]$ we obtain

$$\mathfrak{H}[f, g] = \sum_{i=1}^{\infty} \mathfrak{H}[f, u_i]\mathfrak{H}[g, u_i],$$

Green's formula, together with (23a) and $L[u_i] = -\lambda_i u_i$ immediately yields

$$\mathfrak{H}[g, u_i] = \mathfrak{H}[f, L[u_i]] = -\lambda_i \mathfrak{H}[f, u_i].$$

Thus (23b) is proved, since $\mathfrak{H}[f, u_i] = c_i$.

If $L[f] = g$ does not satisfy the above assumption we obtain the result by a suitable approximation of f; the method is similar to that used to prove (23).

2. The Expansion Theorem. In the case of a single independent variable it is not difficult to derive from our result the theorems on the expansion of arbitrary functions in terms of eigenfunctions, thus supplementing the completeness theorem. This can be done under essentially weaker assumptions than those in Chapter V.

We shall prove: *Every function $f(x)$ which satisfies the admissibility conditions of an eigenvalue variational problem may be expanded in an absolutely and uniformly convergent series $\sum_{n=1}^{\infty} c_n u_n$ in terms of the eigenfunctions.*

Because of the completeness of the orthonormal system $\sqrt{\rho}\,u_n$, it is sufficient to prove that the series $\sum_{n=1}^{\infty} c_n u_n$, where $c_n = \int_0^{\pi} \rho f u_n dx$, converges uniformly (cf. Ch. II, p. 54). To prove this, we again consider the function $\rho_n = f - \sum_{\nu=1}^{n} c_\nu u_\nu$. As above, page 424, we see that

$$\mathfrak{D}[\rho_n] = \mathfrak{D}[f] - \sum_{\nu=1}^{n} c_\nu^2 \lambda_\nu .$$

For sufficiently large n, say $n \geq N$, we have $\lambda_{n+1} \geq 0$, and thus $\mathfrak{D}[\rho_n] \geq 0$. Therefore the series $\sum_{\nu=1}^{\infty} c_\nu^2 \lambda_\nu$ converges, since its terms are non-negative for $\nu > N$. Schwarz's inequality yields

$$\left(\sum_{n=h}^{k} c_n u_n(x)\right)^2 \leq \sum_{n=h}^{k} c_n^2 \lambda_n \sum_{n=h}^{k} \frac{u_n^2(x)}{\lambda_n} \leq \sum_{n=1}^{\infty} c_n^2 \lambda_n \sum_{n=h}^{\infty} \frac{u_n^2(x)}{\lambda_n} .$$

Now, from Ch. V, §11, 3 we know that $| u_n(x) | < C$, where C is a constant independent of n. Since, by §2, 2 and 3, λ_n/n^2 lies between positive bounds, and since $\sum_{n=1}^{\infty} 1/n^2$ converges, $\sum_{n=h}^{k} u_n^2(x)/\lambda_n$ is arbitrarily small uniformly in x for sufficiently large h and k, and thus the same is true of $\sum_{n=h}^{k} | c_n u_n(x) |$. This means that the above series converges absolutely and uniformly, which completes the proof of the expansion theorem.

Our considerations and results remain valid even when singularities occur, as in the case of Legendre and Bessel eigenfunctions. However, our proof of the expansion theorem then holds only if we exclude from the domain an arbitrarily small neighborhood of the singular points, since boundedness of the normalized eigenfunctions has not been proved for such a neighborhood.

3. Generalization of the Expansion Theorem. The asymptotic expressions for the Sturm-Liouville eigenfunctions (Ch. V, §11, 5) permit us to generalize the expansion theorem just proved. In fact, we can prove the theorem: *every piecewise continuous function defined in the fundamental domain with a square-integrable*[1] *first derivative may*

[1] We say that the derivative is square-integrable if the integral of the square of the derivative is bounded for all the intervals of the fundamental domain in which the function is continuous.

be expanded in an eigenfunction series which converges absolutely and uniformly in all subdomains free of points of discontinuity; at the points of discontinuity it represents (like the Fourier series) the arithmetic mean of the right- and left-hand limits. (It should be remarked that this theorem does *not* require that the functions to be expanded satisfy the boundary conditions.)

As in §2, 3, we suppose the differential equation to be written in the form (18),

$$z'' - rz + \lambda z = 0,$$

for the function $z(t)$ in the interval $0 \leq t \leq l$. Then we consider the series

$$G(t, \tau) = \sum_{n=1}^{\infty} \frac{z_n(t) z_n'(\tau)}{\lambda_n}$$

in which z_n denotes the n-th eigenfunction of the above differential equation, say for the boundary condition $z = 0$.

Application of the asymptotic formulas (70) and (71) of Chapter V, together with formula (19), yields

$$G(t, \tau) = \frac{2}{\pi} \sum_{n=1}^{\infty} \frac{\sin n \frac{\pi}{l} t \cos n \frac{\pi}{l} \tau}{n} + \sum_{n=1}^{\infty} \psi_n(t, \tau),$$

where $\psi_n(t, \tau) = O(1/n^2)$, so that $G(t, \tau)$ differs only by an absolutely and uniformly convergent series from

$$\begin{aligned} G^*(t, \tau) &= \frac{2}{\pi} \sum_{n=1}^{\infty} \frac{\sin n \frac{\pi}{l} t \cos n \frac{\pi}{l} \tau}{n} \\ &= \frac{1}{\pi} \sum_{n=1}^{\infty} \frac{1}{n} \left(\sin n \frac{\pi}{l} (t + \tau) + \sin n \frac{\pi}{l} (t - \tau) \right). \end{aligned}$$

As we saw in Ch. II, §5, 1, this series has the following properties: For fixed τ $(0 < \tau \leq l)$ it converges absolutely and uniformly for all t of a closed interval which satisfy the conditions $|t + \tau| > \epsilon$, $|t - \tau| > \epsilon$ for $\epsilon > 0$. Since $\tau > 0$ and $t > 0$, this means that the interval cannot contain the point $t = \tau$. Therefore, although for $t \neq \tau$ the series represents a continuous function, its sum has a finite jump at $t = \tau$ and at this point, again by Ch. II, §5, it is equal to the arithmetic mean of the right- and left-hand limits.

In the case of an arbitrary function which satisfies the above conditions, we remove the discontinuities by adding a suitable sum

$$\sum_i a_i G(t, \tau_i)$$

which is chosen (if necessary) in such a way that the boundary condition is satisfied. We obtain a function satisfying the conditions of the general expansion theorem already proved in subsection 2; this function may, therefore, be expanded in an absolutely and uniformly convergent eigenfunction series. However, as we have just seen, the additional sum may itself be represented by an eigenfunction series with the properties stated in the generalized expansion theorem (see page 427); thus we have proved this theorem for the expansion in terms of the eigenfunctions of the differential equation (18). If we transform the variables z, t back to y, x and the differential equation to the general Sturm-Liouville form, we immediately obtain the expansion theorem for the eigenfunctions $y_n(x)$ of the original differential equation; for, except for constant factors, these eigenfunctions are obtained by multiplying the eigenfunctions z_n by the same nowhere-vanishing function.

§4. *Asymptotic Distribution of Eigenvalues*

The method of §2, 2 and §2, 3 for one independent variable can also be used in the study of the asymptotic behavior of the n-th eigenvalue for several independent variables. We obtain a result significant for physical problems: *the asymptotic behavior of the eigenvalues for differential equations with constant coefficients does not depend on the shape but only on the size of the fundamental domain.*

1. The Equation $\Delta u + \lambda u = 0$ for a Rectangle. In the case of a rectangle of sides a and b, the eigenfunctions and eigenvalues of $\Delta u + \lambda u = 0$ are known explicitly (see Ch. V, §5, 4). In fact, for the boundary condition $u = 0$ they are given—up to a normalizing factor—by the expressions

$$\sin \frac{l\pi x}{a} \sin \frac{m\pi y}{b}, \qquad \pi^2\left(\frac{l^2}{a^2} + \frac{m^2}{b^2}\right) \qquad (l, m = 1, 2, 3, \cdots),$$

and for the boundary condition $\partial u/\partial n = 0$ by the expressions

$$\cos \frac{l\pi x}{a} \cos \frac{m\pi y}{b}, \qquad \pi^2\left(\frac{l^2}{a^2} + \frac{m^2}{b^2}\right) \qquad (l, m = 0, 1, 2, 3, \cdots).$$

If the number of eigenvalues less than a bound λ is denoted in the first case by $A(\lambda)$ and in the second case by $B(\lambda)$, then $A(\lambda)$ and $B(\lambda)$ are equal to the number of integral solutions of the inequality

$$\frac{l^2}{a^2} + \frac{m^2}{b^2} \leq \frac{\lambda}{\pi^2};$$

here $l > 0$, $m > 0$ for the boundary condition $u = 0$ and $l \geq 0$, $m \geq 0$ for the condition $\partial u/\partial n = 0$. Simple *asymptotic expressions* for the desired numbers $A(\lambda)$ and $B(\lambda)$ may now be derived for large λ. $B(\lambda)$, for example, is precisely equal to the number of lattice points with integral coordinates in the sector of the ellipse

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = \frac{\lambda}{\pi^2}$$

which lies in the quadrant $x \geq 0$, $y \geq 0$. For sufficiently large λ the ratio of the area of this sector to the number of lattice points contained in it is arbitrarily close to 1. If a unit square lying above and to the right of each lattice point is associated with it, then the region formed by these squares contains this sector of the ellipse; however, if we omit the squares through which the ellipse passes,—let the number of these be $R(\lambda)$—then the region which remains is contained entirely within this sector of the ellipse. We therefore have between the areas the inequality

$$B(\lambda) - R(\lambda) \leq \lambda \frac{ab}{4\pi} \leq B(\lambda).$$

The arc of the ellipse contained in two adjoining boundary squares has, for sufficiently large λ, at least the length 1. Hence $R(\lambda) - 1$ is at most equal to twice the arc length of the quarter ellipse, and this increases only with $\sqrt{\lambda}$. We thus obtain the asymptotic formula

$$\lim_{\lambda \to \infty} \frac{B(\lambda)}{\lambda} = \frac{ab}{4\pi}$$

or

$$B(\lambda) \sim \lambda \frac{ab}{4\pi}.$$

More precisely, we can write

$$B(\lambda) = \frac{ab}{4\pi}\lambda + \theta c\sqrt{\lambda},$$

where c is a constant independent of λ, and $|\theta| < 1$. This formula is valid for both boundary conditions considered; i.e., it also holds for $A(\lambda)$, since the number of lattice points lying on the line-segments in the boundary of the sector of the ellipse is asymptotically equal to $(a + b)\sqrt{\lambda}/\pi$. If the eigenvalues are written in a sequence $\lambda_1, \lambda_2, \cdots$ ordered according to magnitude, we can calculate the n-th eigenvalue asymptotically by setting $A(\lambda_n) = n$ and $B(\lambda_n) = n$. We obtain

$$\lambda_n \sim \frac{4\pi}{ab} A(\lambda_n) \sim \frac{4\pi}{ab} n$$

or

$$\lim_{n\to\infty} \frac{\lambda_n}{n} = \frac{4\pi}{ab}.$$

2. The Equation $\Delta u + \lambda u = 0$ for Domains Consisting of a Finite Number of Squares or Cubes. Now we consider the equation $\Delta u + \lambda u = 0$ for a domain G which may be decomposed into a finite number, say h, of squares $Q_1, Q_2, \cdots, Q_h$ (or cubes in the case of three independent variables) of side a. Such domains will be called square-domains (or cube-domains). The area of G is then $f = ha^2$ (or its volume is $V = ha^3$).

As before θ will denote a number between -1 and $+1$, and c or C a positive constant. When no misunderstanding seems likely we shall denote different values of θ and c or C by the same symbol without explicitly distinguishing between them by means of indices.

For a domain G consisting of h squares, let $A(\lambda)$ be the number of eigenvalues less than a bound λ for the boundary condition $u = 0$ and let $B(\lambda)$ be the corresponding number for the boundary condition $\partial u/\partial n = 0$. Denoting by $A_{Q_1}(\lambda), A_{Q_2}(\lambda), \cdots, A_{Q_h}(\lambda)$ the corresponding numbers for the subsquares with the boundary conditions $u = 0$, and by $B_{Q_1}(\lambda), B_{Q_2}(\lambda), \cdots, B_{Q_h}(\lambda)$ these numbers with the boundary conditions $\partial u/\partial n = 0$, we have from subsection 1

$$A_{Q_i}(\lambda) = \frac{a^2}{4\pi}\lambda + \theta ca\sqrt{\lambda}, \qquad B_{Q_i}(\lambda) = \frac{a^2}{4\pi}\lambda + \theta' ca\sqrt{\lambda} \tag{26}$$

$$(i = 1, 2, \cdots, h).$$

Combining Theorem 5 with Theorems 2 and 4 (see §2) we have

$$A_{Q_1}(\lambda) + \cdots + A_{Q_h}(\lambda) \leq A(\lambda) \leq B_{Q_1}(\lambda) + \cdots + B_{Q_h}(\lambda).$$

Since the numbers $A_{Q_i}(\lambda)$, $B_{Q_i}(\lambda)$ have the form given by equations (26), we conclude

$$A(\lambda) = \frac{f}{4\pi}\lambda + \theta c a\sqrt{\lambda}.$$

In other words, the following theorem holds:

THEOREM 12: *For all boundary conditions considered, the number $A(\lambda)$ of eigenvalues less than a bound λ of the differential equation $\Delta u + \lambda u = 0$ for a square-domain of area f is asymptotically equal to $f\lambda/4\pi$; that is,*

$$\lim_{\lambda\to\infty} \frac{A(\lambda)}{f\lambda} = \frac{1}{4\pi}. \tag{27}$$

More precisely, for all sufficiently large λ the relation

$$\left| \frac{4\pi A(\lambda)}{f\lambda} - 1 \right| < \frac{C}{\sqrt{\lambda}} \tag{28}$$

holds, where C is a constant independent of λ.

If ρ_n denotes the n-th eigenvalue corresponding to one of the boundary conditions in question, this theorem or relation (28) is equivalent to equation

$$\rho_n = \frac{4\pi}{f} n + \theta c\sqrt{n}, \tag{29}$$

where again $-1 \leq \theta \leq 1$ and c is a constant independent of n. To see this one has only to set $A(\rho_n) = n$ in (28).

Theorem 12 remains valid even if the function σ in the boundary condition $\partial u/\partial n + \sigma u = 0$ assumes negative values. This will again be shown with the aid of the remarks of §2, 5. First we note that according to Theorem 5, the n-th eigenvalue μ_n for the boundary condition $\partial u/\partial n + \sigma u = 0$ can surely be no greater than the n-th eigenvalue λ_n for the boundary condition $u = 0$. Thus we may assume from the outset that the expression

$$\mathfrak{D}[\varphi] = D[\varphi] + \int_\Gamma p\sigma\varphi^2\, ds,$$

whose maximum-minimum is clearly μ_n, never exceeds the bound λ_n for any of the functions φ which are admissible in the variational problem with the boundary condition $u = 0$.

Now by §2, 5 we have

$$\left| \int_\Gamma p\sigma\varphi^2 \, ds \right| < c_1 \sqrt{\,|\, D[\varphi] \,|} + c_2 ,$$

where c_1, c_2 are numerical constants; thus we have

$$D[\varphi] - c_1 \sqrt{\,|\, D[\varphi] \,|} - c_2 < \mathfrak{D}[\varphi] < D[\varphi] + c_1 \sqrt{\,|\, D[\varphi] \,|} + c_2 .$$

From the assumption $\mathfrak{D}[\varphi] \leq \lambda_n$ it follows that

$$D[\varphi] - c_1 \sqrt{\,|\, D[\varphi] \,|} - c_2 < \lambda_n ,$$

and this in turn implies that for increasing n, $D[\varphi]$ can increase no faster than λ_n, or in other words that there must be a relation of the form

$$D[\varphi] < c_3 \lambda_n ,$$

where c_3 also denotes a constant. Since relation (29) holds for $\rho_n = \lambda_n$ we have, under the assumptions made with respect to φ,

$$D[\varphi] - c_4 \sqrt{n} \leq \mathfrak{D}[\varphi] \leq D[\varphi] + c_4 \sqrt{n};$$

this relation is valid for the lower bounds of the expressions $\mathfrak{D}[\varphi]$ for given functions v_1, v_2, $\cdots$, v_{n-1}, and hence also for the maxima of these lower bounds. The maximum-minimum of $D[\varphi]$ is the n-th eigenvalue for the boundary condition $\partial u/\partial n = 0$, for which relation (29) has already been proved. Therefore (29) follows directly for the maximum of the lower bounds of $\mathfrak{D}[\varphi]$, i.e. for the n-th eigenvalue μ_n with the boundary condition $\partial u/\partial n + \sigma u = 0$; this relation is equivalent to the assertion of Theorem 12.

If there are three independent variables instead of two, the preceding discussion remains unchanged except for the expressions A_{Q_i} and B_{Q_i} for the number of eigenvalues less than the bound λ under the boundary conditions $u = 0$ and $\partial u/\partial n = 0$, respectively. We find, in fact, that

$$\text{(26a)} \quad A_{Q_i}(\lambda) = \frac{1}{6\pi^2} a^3 \lambda^{3/2} + \theta c a^2 \lambda, \qquad B_{Q_i}(\lambda) = \frac{1}{6\pi^2} a^3 \lambda^{3/2} + \theta c a^2 \lambda$$

holds and obtain

THEOREM 13: *Consider the differential equation* $\Delta u + \lambda u = 0$ *for a polyhedron of volume* V *consisting of a finite number* k *of congruent cubes. For all boundary conditions in question the number* $A(\lambda)$ *of eigenvalues less than a bound* λ *is asymptotically equal to* $V\lambda^{3/2}/6\pi^2$, *i.e.*

$$\lim_{\lambda\to\infty} \frac{A(\lambda)}{V\lambda^{3/2}} = \frac{1}{6\pi^2}. \tag{27a}$$

More precisely, for sufficiently large λ *we have the formula*

$$\left| \frac{6\pi^2 A(\lambda)}{V\lambda^{3/2}} - 1 \right| < C\,\frac{1}{\sqrt{\lambda}}, \tag{28a}$$

in which C *is a constant independent of* λ.[1]

3. Extension to the General Differential Equation $L[u] + \lambda\rho u = 0$. We shall now extend Theorem 13 to the general self-adjoint differential equation (1). We assume that, by repeated halving of the length of the edge a, the subdivision of the domain G into squares or cubes has been refined in such a way that in the resulting subdomains the difference between the largest and the smallest values of the functions p and ρ is always less than a given small positive number ϵ. *The function* q *can have no influence at all on the asymptotic distribution of the eigenvalues,* since the expression $\mathfrak{D}[\varphi]$, and with it its maximum-minimum, changes by an amount which is bounded, namely, by less than $|q_M|/\rho_m$, here q_M and ρ_m have the same meaning as before. Accordingly, we set $q = 0$ without affecting the asymptotic distribution.

We consider the case of a plane domain G consisting of a finite number of squares. Let the number of squares again be h, and let the lengths of their sides be a. $A'(\lambda)$ will denote the number of eigenvalues less than a bound λ of the differential equation $L[u] + \lambda\rho u = 0$ for the domain G; any of the boundary conditions considered may be used, but for the condition $\partial u/\partial n + \sigma u = 0$ we must make the restrictive assumption $\sigma \geq 0$. We denote the subsquares by $Q_1, Q_2, \cdots, Q_h$, and the associated numbers of eigenvalues of the differential equation less than a bound λ by $A'_{Q_1}(\lambda), A'_{Q_2}(\lambda), \cdots, A'_{Q_h}(\lambda)$ for the boundary condition $u = 0$, and by $B'_{Q_1}(\lambda), B'_{Q_2}(\lambda), \cdots, B'_{Q_h}(\lambda)$ for the boundary condition $\partial u/\partial n = 0$. From Theorems 2, 4 and 5 we obtain

[1] It is not in general possible to obtain a sharper estimate of the error in this asymptotic estimate of $A(\lambda)$, since for both the square and the cube the indicated order of magnitude of the error is actually attained.

(30) $A'_{Q_1}(\lambda) + \cdots + A'_{Q_h}(\lambda) \leq A'(\lambda) \leq B'_{Q_1}(\lambda) + \cdots + B'_{Q_h}(\lambda).$

Theorem 7 implies

$$A'_{Q_i}(\lambda) \geq \frac{\rho_m^{(i)}}{p_M^{(i)}} A_{Q_i}(\lambda), \qquad B'_{Q_i}(\lambda) \leq \frac{\rho_M^{(i)}}{p_m^{(i)}} B_{Q_i}(\lambda)$$

where by $p_M^{(i)}$, $\rho_M^{(i)}$ we denote the maxima and by $p_m^{(i)}$, $\rho_m^{(i)}$ the minima of the respective functions in the square Q_i; again (cf. subsection 2) the number of corresponding eigenvalues of the differential equation $\Delta u + \lambda u = 0$ is given by equations (26) and denoted by $A_{Q_i}(\lambda)$ or $B_{Q_i}(\lambda)$. For, substituting $p_M^{(i)}$ for p and $\rho_M^{(i)}$ for ρ in differential equation (1), we see by Theorem 7 that each eigenvalue becomes larger (or at least is not diminished), and thus the number of eigenvalues less than a fixed λ decreases (or does not increase). On the other hand, differential equation (1) goes over into the differential equation

$$\Delta u + \lambda \frac{\rho_m^{(i)}}{p_M^{(i)}} u = 0,$$

whose eigenvalues are the eigenvalues of $\Delta u + \lambda u = 0$ multiplied by the factor $p_M^{(i)}/\rho_m^{(i)}$. A corresponding argument holds if $p_m^{(i)}$ is substituted for p and $\rho_M^{(i)}$ for ρ.

Furthermore, since ρ and p are continuous functions we have

$$\iint_G \frac{\rho}{p}\, dx\, dy = a^2 \sum_{i=1}^{h} \frac{\rho_m^{(i)}}{p_M^{(i)}} + \delta = a^2 \sum_{i=1}^{h} \frac{\rho_M^{(i)}}{p_m^{(i)}} + \delta',$$

where the numbers $|\delta|$, $|\delta'|$ may be made arbitrarily small by making the original subdivision into squares sufficiently fine, that is, by taking a sufficiently small. As in subsection 2, we apply (30) and find

$$A(\lambda) = \frac{\lambda}{4\pi} \iint_G \frac{\rho}{p}\, dx\, dy + \lambda\delta'' + \theta c\sqrt{\lambda},$$

where $|\delta''|$ also is arbitrarily small. This is equivalent to the following statement about the asymptotic distribution of eigenvalues:

THEOREM 14: *For the differential equation* $L[u] + \lambda\rho u = 0$, *under any of the boundary conditions considered, the number* $A(\lambda)$ *of eigenvalues less than a given bound* λ *for a square-domain* G *is asymptotically equal to*

$$\frac{\lambda}{4\pi} \iint_G \frac{\rho}{p}\, dx\, dy;$$

in other words the relation

$$\lim_{\lambda\to\infty} \frac{A(\lambda)}{\lambda} = \frac{1}{4\pi} \iint_G \frac{\rho}{p}\, dx\, dy \tag{31}$$

holds.

As in subsection 2, we see that the original assumption $\sigma \geq 0$ is unnecessary.

Analogous considerations for three-dimensional space lead to

THEOREM 15: *For the differential equation* $L[u] + \lambda\rho u = 0$, *under any of the boundary conditions considered, the number of eigenvalues less than a given bound* λ *for a cube-domain* G *is asymptotically equal to*

$$\frac{1}{6\pi^2} \lambda^{3/2} \iiint_G \left(\frac{\rho}{p}\right)^{3/2} dx\, dy\, dz;$$

in other words, the relation

$$\lim_{\lambda\to\infty} \frac{A(\lambda)}{\lambda^{3/2}} = \frac{1}{6\pi^2} \iiint_G \left(\frac{\rho}{p}\right)^{3/2} dx\, dy\, dz \tag{32}$$

holds.

Finally, we mention that the arguments of the two preceding subsections may be applied also to a *more general domain* consisting of a finite number of *rectangles or rectangular parallelepipeds.*

4. Asymptotic Distribution of Eigenvalues for an Arbitrary Domain. With the boundary condition $u = 0$, the same asymptotic formulas can easily be proved for an arbitrary domain G which can be approximated from within and without by special regions consisting only of a finite number of squares and differing arbitrarily little in area.

However, we shall aim to establish at once the asymptotic laws for arbitrary G and all the boundary conditions under consideration; for this purpose a somewhat deeper analysis is required.

First we consider a plane domain G whose boundary has continuous curvature and restrict ourselves to the differential equation $\Delta u + \lambda u = 0$.

We examine briefly the eigenvalues associated with this differential equation under the boundary condition $\partial u/\partial n = 0$, investigating the number of eigenvalues less than a given bound for a few simple domains.

Let G be an isosceles right triangle with sides of length a. Every eigenfunction for the triangle is also an eigenfunction for the square which is obtained by reflection across the hypotenuse under the same

boundary condition $\partial u/\partial n = 0$. For, it is immediately clear that any eigenfunction may be continued into the reflected triangle if the functional value of each original point is assigned to its reflection; in this way the boundary condition $\partial u/\partial n = 0$ is satisfied along the entire boundary of the square. The n-th eigenvalue for the triangle is therefore also an eigenvalue for the square; hence the n-th eigenvalue for the square is certainly no larger than that for the triangle. In other words, *for the triangle, with the boundary condition* $\partial u/\partial n = 0$, *the number of eigenvalues less than a given bound is at most equal to the corresponding number for the square,* that is, the number given by formula (26).

Next, let G be an arbitrary right triangle of sides a and b, where we assume $b \leq a$. Let side a lie on the x-axis, side b on the y-axis. By means of the transformation $\xi = x$, $\eta = ay/b$ we transform the triangle G into an isosceles right triangle G' with sides equal to a. The expression $D[\varphi]$ then becomes

$$D[\varphi] = \iint_{G'} \left[\left(\frac{\partial \varphi}{\partial \xi} \right)^2 + \frac{a^2}{b^2} \left(\frac{\partial \varphi}{\partial \eta} \right)^2 \right] \frac{b}{a} \, d\xi \, d\eta$$

and the auxiliary condition $H[\varphi] = 1$ becomes

$$\iint_{G'} \varphi^2 \frac{b}{a} \, d\xi \, d\eta = 1,$$

while the additional auxiliary conditions $H[\varphi, v_i] = 0$ of §1, 4 retain their form under this transformation. Thus, if we omit the insignificant constant factor b/a occurring in both integrals, we can characterize the n-th eigenvalue κ_n for the triangle as the maximum-minimum of the integral

$$\iint_{G'} \left[\left(\frac{\partial \varphi}{\partial \xi} \right)^2 + \frac{a^2}{b^2} \left(\frac{\partial \varphi}{\partial \eta} \right)^2 \right] d\xi \, d\eta$$

taken over G', where the maximum-minimum is to be understood in the usual sense. Since $a/b \geq 1$, the relation

$$\iint_{G'} \left[\left(\frac{\partial \varphi}{\partial \xi} \right)^2 + \frac{a^2}{b^2} \left(\frac{\partial \varphi}{\partial \eta} \right)^2 \right] d\xi \, d\eta \geq \iint_{G'} \left[\left(\frac{\partial \varphi}{\partial \xi} \right)^2 + \left(\frac{\partial \varphi}{\partial \eta} \right)^2 \right] d\xi \, d\eta$$

holds, and the maximum-minimum of the left side is therefore at least as large as that of the right side, i.e. it is at least as large as the n-th eigenvalue for the isosceles right triangle G', and thus a fortiori

larger than the n-th eigenvalue of a square of side a. Therefore, *under the boundary condition* $\partial u/\partial n = 0$, *the number of eigenvalues less than a given bound for a right triangle with sides not larger than a is certainly no larger than the corresponding number of eigenvalues for the square of side a; hence it is no larger than the corresponding number for any larger square.*

Similarly, *the number of eigenvalues less than a given bound for an arbitrary rectangle is never larger than the corresponding number for a square whose side is at least equal to the larger side of the rectangle.*

These results, together with Theorem 4, show that it is possible to obtain an upper bound for the number of eigenvalues less than a given bound whenever the domain in question is composed of a finite number of rectangles and right triangles.

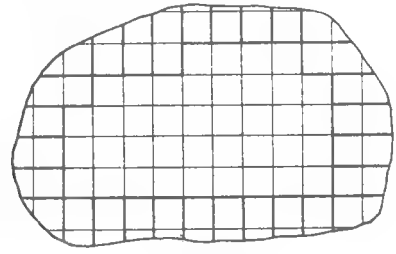


Figure 4

We now consider how the distribution of the eigenvalues is affected by the boundary strip which remains when we approximate the domain G by means of squares. First we define the boundary strip: Let the subdivision into squares be so fine that, for every segment of the boundary of G which is contained in one of the squares, the direction of the normal varies by less than a given small angle η, whose size will be determined later. (This can be accomplished by repeated halving of the side of the square.) Then, as in Figure 4, we can associate with the boundary Γ a number r of adjacent elementary domains $E_1, E_2, \cdots, E_r$ in the following manner: Each domain E_i is bounded either by two perpendicular line-segments AB, AC of the partition, whose lengths lie between a and $3a$, and a segment BC of the boundary (Figure 5), or by a line-segment AB of the partition, two line-segments, AC, BD of lengths between a and $3a$ perpendicular to AB, and a segment CD of the boundary (Figure 6). We construct a bound-

ary strip S consisting of r domains of this kind, such that when this strip is removed from G a square-domain is left, consisting of h squares $Q_1, Q_2, \cdots, Q_h$.[1] The number r is evidently smaller than a constant C/a, where C is independent of a and depends essentially on the length of the boundary.

In order to obtain an upper bound for the number $B_{E_i}(\lambda)$ of eigenvalues less than some bound λ of the differential equation $\Delta u + \lambda u = 0$ for the domain E_i with boundary condition $\partial u/\partial n = 0$, we must again try to find a lower bound for the n-th eigenvalues. To this end, we take any point on the curvilinear boundary segment of E_i and draw the tangent through it. This tangent together with the straight boundary segments of E_i forms a domain of the type $AB'C'$ (Fig. 5),

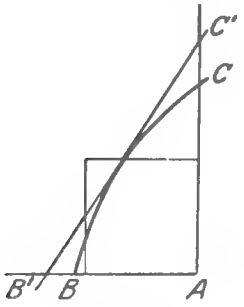


Figure 5

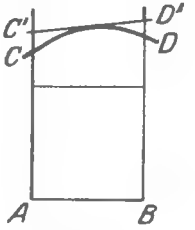


Figure 6

e.g. if η is sufficiently small it forms a right triangle with sides smaller than $4a$, or else a trapezoid of the type $ABC'D'$ (Figure 6) the sides AC', BD' of which are also smaller than $4a$; the shape of the resulting domain depends on the type to which E_i belongs. We shall denote the domains $AB'C'$ and $ABC'D'$ by E_i'. The domain E_i can always be deformed into the domain E_i' by a transformation of the form (21), as discussed in §2. In the case of domains of the first

[1] We leave it to the reader to carry out this construction: Divide the boundary curve into a finite number of arcs which are of three kinds. On arcs of the first kind the tangents should form an angle of at most 30° with the x-axis, on those of the second kind one of at most 30° with the y-axis; on arcs of the third kind the tangents should not form angles with either axis of less than 20°. The end points of arcs of the first and second kind should have rational abscissæ and ordinates, respectively. If the subdivision into squares (on whose sides these end points lie) is sufficiently fine the construction is possible.

type, let A be the pole of a system of polar coordinates ρ, θ, and let $\rho = f(\theta)$ be the equation of the curved segment BC, $\rho = g(\theta)$ the equation of the line $B'C'$. Then the equations

$$\theta' = \theta, \qquad \rho' = \rho \frac{g(\theta)}{f(\theta)}$$

represent a transformation of the curvilinear triangle E_i into the rectilinear triangle E_i'. For a domain of the second type $ABCD$, let AB lie along the x-axis, let $y = g(x)$ be the equation of the line-segment $C'D'$ and let $y = f(x)$ be the equation of the curved segment CD. We then consider the transformation

$$x' = x, \qquad y' = y \frac{g(x)}{f(x)}.$$

If we assume that the fundamental interval a is sufficiently small, and therefore that the total rotation of the tangent on the curved segment CB or CD is taken sufficiently small, then the transformations considered here evidently have precisely the form (21), and the quantity denoted by ϵ in (21) is arbitrarily small. It follows from the corollary to Theorem 10 that the corresponding n-th eigenvalues for the domains E_i and E_i' differ only by a factor which itself differs by a small amount from 1, uniformly for all n. Hence the same is true also for the corresponding numbers $B_{E_i}(\lambda)$ and $B_{E_i'}(\lambda)$ of eigenvalues less than the bound λ for the boundary condition $\partial u/\partial n = 0$.

The domain E_i' is either a right triangle with sides smaller than $4a$ or a combination of such a triangle and a rectangle with sides smaller than $3a$; it follows that if a is taken sufficiently small, the number $B_{E_i}(\lambda)$ from some λ on satisfies the inequality

$$B_{E_i}(\lambda) < c_1 a^2 \lambda + c_2 a\sqrt{\lambda} \tag{33}$$

where c_1, c_2 are numerical constants, to be chosen suitably.

We can now deduce the laws that govern the asymptotic distribution of eigenvalues for the domain G. Let $A(\lambda)$, for any of the boundary conditions considered, again denote the number of eigenvalues less than a bound λ of the differential equation $\Delta u + \lambda u = 0$ for the domain G, where again, if necessary, we make the assumption $\sigma \geq 0$. Suppose the plane is partitioned into squares of side a, inducing a decomposition of the domain G into h squares $Q_1, Q_2, \cdots, Q_h$ and r boundary domains $E_1, E_2, \cdots, E_r$. For the square Q_i the number of

eigenvalues less than λ is again denoted by $A_i(\lambda)$ for the boundary condition $u = 0$ and by $B_i(\lambda)$ for the condition $\partial u/\partial n = 0$. The corresponding numbers for the domains E_i are denoted by $A_{E_i}(\lambda)$ and $B_{E_i}(\lambda)$, respectively (we shall use only the numbers $B_{E_i}(\lambda)$).

From equations (26) we have

$$A_i(\lambda) = \frac{a^2}{4\pi}\lambda + a\theta_1 c_1\sqrt{\lambda}, \qquad B_i(\lambda) = \frac{a^2}{4\pi}\lambda + a\theta_2 c_2\sqrt{\lambda}$$

and, by (33),

$$B_{E_i}(\lambda) = \theta_3(c_3\lambda a^2 + ac_4\sqrt{\lambda}),$$

where, as always, θ_1, θ_2, θ_3 denote numbers between -1 and $+1$, and c_1, c_2, c_3, c_4 are constants independent of a, i, and λ.

From Theorems 5, 2, and 4 we find

$$A_1(\lambda) + A_2(\lambda) + \cdots + A_h(\lambda) \leq A(\lambda)$$
$$\leq B_1(\lambda) + \cdots + B_h(\lambda) + B_{E_1}(\lambda) + \cdots + B_{E_r}(\lambda);$$

furthermore, we have

$$A_1(\lambda) + \cdots + A_h(\lambda) = \frac{ha^2}{4\pi}\lambda + \theta_1 c_1 ha\sqrt{\lambda} = \lambda\left(\frac{ha^2}{4\pi} + \frac{\theta_1 c_1 ha}{\sqrt{\lambda}}\right),$$

$$B_1(\lambda) + \cdots + B_h(\lambda) + B_{E_1}(\lambda) + \cdots + B_{E_r}(\lambda)$$
$$= \frac{ha^2}{4\pi}\lambda + \theta_2 c_2 ha\sqrt{\lambda} + \theta_3 ra^2\lambda c_3 + \theta_3 rac_4\sqrt{\lambda}$$
$$= \lambda\left[\left(\frac{ha^2}{4\pi} + \theta_3 c_3 ra^2\right) + (ha\theta_2 c_2 + ra\theta_3 c_4)\frac{1}{\sqrt{\lambda}}\right].$$

Now $ar < c_5$; therefore, for sufficiently small a, a^2r is arbitrarily small and we have

$$|ha^2 - f| < \delta,$$

no matter how small δ is. From these inequalities the asymptotic relation

$$\lim_{\lambda\to\infty}\frac{4\pi A(\lambda)}{\lambda f} = 1$$

follows immediately. For, we may choose the quantity a arbitrarily, and by taking a sufficiently small fixed a, make the factors of λ in

the above inequalities arbitrarily close to the value $f/4\pi$ for sufficiently large λ.

Even without the assumption $\sigma \geq 0$, we obtain the same asymptotic law by means of the arguments used at an analogous point in subsection 2. Summing up, we have

THEOREM 16: *Under any of the boundary conditions considered, the number $A(\lambda)$ of eigenvalues less than a bound λ of the differential equation $\Delta u + \lambda u = 0$ for the domain G is asymptotically equal to $\lambda f/4\pi$; in other words,*

$$\lim_{\lambda\to\infty} \frac{4\pi A(\lambda)}{\lambda f} = 1, \tag{34}$$

where f denotes the area of the domain.

In the proof, we first made the assumption that the boundary Γ of G had no corners. However, both the result and the argument remain essentially unchanged if a finite number of vertices is admitted.

The preceding argument remains valid if we deal with the more general equation $L[u] + \lambda\rho u = 0$ instead of the differential equation $\Delta u + \lambda u = 0$. As in subsection 3, we obtain

THEOREM 17: *Under any of the boundary conditions considered, the number $A(\lambda)$ of eigenvalues less than a fixed λ of the differential equation $L[u] + \lambda\rho u = 0$ for G is asymptotically equal to $\frac{\lambda}{4\pi}\iint_G \frac{\rho}{p}\,dx\,dy$; in other words,*

$$\lim_{\lambda\to\infty} \frac{A(\lambda)}{\lambda} = \frac{1}{4\pi}\iint_G \frac{\rho}{p}\,dx\,dy.$$

Considerations similar to those given here for the plane lead to the following result for the eigenvalue problem in space:

THEOREM 18: *Under any of the boundary conditions considered, the number $A(\lambda)$ of eigenvalues less than a fixed λ of the equation $\Delta u + \lambda u = 0$ for a space-domain of volume V is asymptotically equal to $\lambda^{3/2}V/6\pi^2$; in other words,*

$$\lim_{\lambda\to\infty} \frac{A(\lambda)}{\lambda^{3/2}V} = \frac{1}{6\pi^2}. \tag{35}$$

THEOREM 19: *The corresponding number for the more general differential equation $L[u] + \lambda\rho u = 0$ is asymptotically equal to $\frac{\lambda^{3/2}}{6\pi^2}\iiint_G \left(\frac{\rho}{p}\right)^{3/2} dx\,dy\,dz$; in other words,*

$$\lim_{\lambda\to\infty}\frac{A(\lambda)}{\lambda^{3/2}}=\frac{1}{6\pi^2}\iiint_G\left(\frac{\rho}{p}\right)^{3/2}dx\,dy\,dz. \tag{36}$$

We assume here that G has a boundary consisting of a finite number of surface elements with continuous curvature, which do not come in contact with each other but may form edges and vertices.

5. Sharper Form of the Laws of Asymptotic Distribution of Eigenvalues for the Differential Equation $\Delta u + \lambda u = 0$. The asymptotic eigenvalue laws can be formulated more precisely by establishing an *estimate of the error which arises* when the expression $A(\lambda)$ is replaced by its asymptotic value; we restrict ourselves to the differential equation $\Delta u + \lambda u = 0$.

We need only approximate the domain G by elementary domains composed of squares or cubes in such a way that these domains are no smaller and no more numerous than necessary. First, let G be a domain in the plane. We construct a sequence of approximating square-domains: We begin with a subdivision of the plane into squares, each, say, of side 1, and suppose that, of these, h_0 squares $Q_1^0, Q_2^0, \cdots, Q_{h_0}^0$ lie entirely in the interior of G. Now let each square be decomposed into four congruent squares each of side $\frac{1}{2}$, and let h_1 of these squares, $Q_1^1, Q_2^1, \cdots, Q_{h_1}^1$, lie in the interior of G but not in the interior of any of the squares Q_i^0. Continuing in this way, after t steps we obtain h_t squares $Q_1^t, Q_2^t, \cdots, Q_{h_t}^t$, each of side $1/2^t$, which lie in the interior of G but in none of the preceding squares. In accordance with the preceding subsection, we proceed in such a way that the boundary strip remaining after the t-th approximation consists of r subdomains $E_1, E_2, \cdots, E_r$ of the type defined there; here the number a is equal to $1/2^t$.

By our assumptions on the boundary, the numbers h_i and r satisfy relations

$$h_i < 2^i c, \qquad r < 2^t c, \tag{37}$$

where c is a constant independent of i and t, determined essentially by the length of the boundary.[1]

Again we denote by $A_m^i(\lambda)$ and $A_{E_m}(\lambda)$ the number of eigenvalues less than a bound λ for the domains Q_m^i and E_m for the boundary condition $u = 0$, and by $B_m^i(\lambda)$ and $B_{E_m}(\lambda)$ the number for the bound-

[1] These inequalities mean that the perimeters of the approximating square-domains and those of the residual boundary strips are not of a higher order of magnitude than the perimeter of G.

ary condition $\partial u/\partial n = 0$. If the function σ in the boundary condition $\partial u/\partial n + \sigma u = 0$ is non-negative, we have by Theorems 2, 4 and 5

$$A(\lambda) \leq (B_1^0 + B_2^0 + \cdots + B_{h_0}^0) + \cdots + (B_1^t + B_2^t + \cdots + B_{h_t}^t)$$

$$+ (B_{E_1} + B_{E_2} + \cdots + B_{E_r}), \tag{38}$$

$$A(\lambda) \geq (A_1^0 + A_2^0 + \cdots + A_{h_0}^0) + \cdots + (A_1^t + A_2^t + \cdots + A_{h_t}^t).$$

By (26) and (33) the right side of the first of these inequalities is equal to

$$\frac{1}{4\pi}\left(h_0 + \frac{h_1}{2^2} + \frac{h_2}{2^4} + \cdots + \frac{h_t}{2^{2t}} + \frac{r\theta c}{2^{2t}}\right)\lambda$$

$$+ \theta_1 c_2\left(h_0 + \frac{h_1}{2} + \frac{h_2}{2^2} + \cdots + \frac{h_t}{2^t} + \frac{r}{2^t}\right)\sqrt{\lambda};$$

because

$$h_0 + \frac{h_1}{2^2} + \cdots + \frac{h_t}{2^{2t}} + \frac{r}{2^{2t}} = f - \theta_2 c_2 \frac{r}{2^{2t}}$$

and (37) holds, this right side has the form

$$\frac{1}{4\pi}\left(f + \frac{c\theta_3 c_3}{2^t}\right)\lambda + \theta_4 c_4 (t+2)\sqrt{\lambda}$$

and we obtain the inequality

$$(B_1^0 + \cdots + B_{h_0}^0) + \cdots + (B_{E_1} + \cdots + B_{E_r})$$

$$\leq \frac{f}{4\pi}\lambda + C\left(\frac{\lambda}{2^t} + t\sqrt{\lambda}\right), \tag{39}$$

which holds for sufficiently large t, where as always C denotes a constant independent of λ and t.

We choose the number t, which is still at our disposal, in such a way that the two terms inside the parentheses are as nearly equal as possible, i.e. $t =$ largest integer $\leq \log\lambda/\log 4$. Then for sufficiently large λ we have from (38) and (39)

$$A(\lambda) \leq \frac{f}{4\pi}\lambda + C\sqrt{\lambda}\log\lambda. \tag{40}$$

Precisely the same form is found for the lower bound of the expression $A(\lambda)$, with negative C.

We assumed up to now that if the function σ occurs in the boundary condition it never becomes negative. However, the argument of subsection 2 shows that in view of inequality (20) of §2, 5 the bounds retain their form even without this restriction. Thus we obtain in general the sharper asymptotic law:

THEOREM 20: *Under any of the boundary conditions considered, the difference* $A(\lambda) - f\lambda/4\pi$ *is, for* $\lambda \to \infty$, *of order not greater than* $\sqrt{\lambda} \log \lambda$.

The same argument, carried out for three-dimensions, leads to

THEOREM 21: *Under any of the boundary conditions considered for the problem of a domain of volume* V *in space, the difference* $A(\lambda) - V\lambda^{3/2}/6\pi^2$ *is, for* $\lambda \to \infty$, *of order not greater than* $\lambda \log \lambda$.

§5. *Eigenvalue Problems of the Schrödinger Type*

In Ch. V, §12 we considered Schrödinger's eigenvalue problem for an infinite fundamental domain and studied the properties of the associated spectrum. We shall now approach this problem from the standpoint of the calculus of variations; this approach is significant although its results are incomplete. Not only in the Schrödinger example, but also in other eigenvalue problems for infinite domains—problems which cannot be solved by the separation of variables—we shall see that the spectrum contains a countably infinite sequence of increasing discrete negative eigenvalues. (It also contains a "continuous spectrum.")

Let the eigenvalue equation be

$$\Delta u + Vu + \lambda u = 0, \tag{41}$$

with the condition that the function $u(x, y, z)$ remain finite at infinity. We suppose that the function $V(x, y, z)$—the negative of the potential energy—is positive throughout the space and vanishes at infinity in accordance with the following inequalities, valid for sufficiently large r:

$$\frac{A}{r^\alpha} < V < \frac{B}{r^\beta}, \tag{42}$$

where A and B are positive constants and the exponents satisfy the relations

$$0 < \beta \leq \alpha < 2.$$

In addition, we permit V to become infinite at the origin,[1] of an order

[1] The following discussion remains applicable as long as V becomes singular in this manner at a finite number of points.

no higher than C/r^γ, where $0 \leq \gamma < 2$; r denotes the distance of the point (x, y, z) from the origin.

If $\int \cdots dg$ stands for integration over the whole x, y, z-space, then with the usual notation the variational problem leading to the eigenvalue λ_n and the eigenfunction u_n may be written as follows:

$$J[\varphi] = \int (\varphi_x^2 + \varphi_y^2 + \varphi_z^2 - V\varphi^2)\, dg = \text{max. min.} \tag{43}$$

with the auxiliary conditions

$$\int \varphi^2\, dg = 1$$

$$\int \varphi v_\nu\, dg = 0 \qquad (\nu = 1, 2, \cdots, n-1). \tag{44}$$

Here $\varphi(x, y, z)$, along with its first derivatives, is assumed to be continuous and square-integrable over the whole space. We assume also that $\int V\varphi^2\, dg$ exists; as before, $v_1, v_2, \cdots, v_{n-1}$ denote piecewise continuous functions.

We prove first that this variational problem has a meaning, or in other words, that under the given conditions the integral $J[\varphi]$ is bounded from below. In order to do this we need only note that

$$V \leq \frac{a}{r^2} + b$$

holds everywhere; by choosing the positive constant b sufficiently large we may take the positive constant a to be arbitrarily small. Hence

$$\int V\varphi^2\, dg \leq a \int \frac{1}{r^2}\varphi^2\, dg + b \int \varphi^2\, dg. \tag{45}$$

We now apply the integral inequality

$$\int \frac{1}{r^2}\varphi^2\, dg \leq 4 \int (\varphi_x^2 + \varphi_y^2 + \varphi_z^2)\, dg, \tag{46}$$

which is proved as follows: The substitution $\psi = \varphi\sqrt{r}$ yields

$$\varphi_x^2 + \varphi_y^2 + \varphi_z^2 = \frac{1}{r}(\psi_x^2 + \psi_y^2 + \psi_z^2) - \frac{1}{r^2}\psi\psi_r + \frac{1}{4r^3}\psi^2,$$

and therefore

$$\int (\varphi_x^2 + \varphi_y^2 + \varphi_z^2)\, dg \geq -\int \frac{1}{2r^2}(\psi^2)_r\, dg + \frac{1}{4}\int \frac{1}{r^3}\psi^2\, dg.$$

The first term on the right may be integrated explicitly; since $\int \varphi^2 dg$ exists by hypothesis, it has the value zero.[1] This yields the desired inequality. With its help we obtain from (45)

$$J[\varphi] \geq (1 - 4a)\int (\varphi_x^2 + \varphi_y^2 + \varphi_z^2)\, dg - b,$$

and if, as we may, we take $a < \frac{1}{4}$, we find

$$J[\varphi] \geq -b,$$

which proves that $J[\varphi]$, and hence also the eigenvalues of (41), are bounded from below.

In order to obtain upper bounds for the eigenvalues, we strengthen the admissibility conditions of our variational problem by requiring in addition that φ vanish identically outside a sphere K_R of radius R with center at the origin. According to our general principles, the n-th eigenvalue $\nu_n(R)$ of the resulting problem for the sphere K_R satisfies the inequality $\nu_n(R) \geq \lambda_n$. On the other hand, it may be estimated easily in terms of the eigenvalue $\mu_n(R)$ of the differential equation $\Delta u + \mu u = 0$ for the sphere K_R with vanishing boundary values. In fact, from assumption (42) we have $V \geq A/R^\alpha$ in K_R (for sufficiently large R) and

$$\int_{K_R} (\varphi_x^2 + \varphi_y^2 + \varphi_z^2 - V\varphi^2)\, dg \leq \int_{K_R} (\varphi_x^2 + \varphi_y^2 + \varphi_z^2)\, dg - \frac{A}{R^\alpha}\int_{K_R} \varphi^2\, dg.$$

Hence it follows immediately that

$$\nu_n(R) \leq \mu_n(R) - \frac{A}{R^\alpha}.$$

But $\mu_n(R) = \mu_n(1)/R^2$, where the $\mu_n(1)$ are the eigenvalues for the unit sphere, and we obtain

$$\lambda_n \leq \frac{\mu_n(1)}{R^2} - \frac{A}{R^\alpha}.$$

[1] In fact, there must exist a sequence of values $R_1, R_2, \cdots, R_n, \cdots$ such that the integrals $\frac{1}{R_n}\int \varphi^2\, dS$—extended over the surfaces of the spheres with the radii R_n—approach zero as R_n becomes infinite. First we integrate over these spheres and then pass to the infinite domain in the limit.

Since $\alpha < 2$, for given n the right side is negative when R is taken sufficiently large.

Thus we have proved that our variational problems yield a sequence of monotonically nondecreasing negative eigenvalues.

To prove that these eigenvalues approach zero as n increases, we may estimate them (see also page 449) in terms of the eigenvalues κ_n of the special Schrödinger problem for $V = c/r$, for which they are known explicitly and for which κ_n tends to zero according to Ch. V, §12, 4. We need only remark that an inequality $V \leq a/r^2 + b/r + k$ holds, where for sufficiently large b the positive constants a and k may be taken arbitrarily small. Thus, using relation (46), we evidently have

$$\lambda_n \geq (1-4a)\kappa_n - k$$

if we take $c = b/(1 - 4a)$. It follows that as n increases the eigenvalue λ_n at some time exceeds the value $-2k$; hence it converges to zero, since k may be taken arbitrarily small.

The occurrence of a continuous spectrum of positive eigenvalues is plausible if we consider the eigenvalue problem for the infinite domain as the limiting case of the eigenvalue problem for a finite domain, e.g. for the sphere K_R with increasing radius R. In fact, the n-th eigenvalue $\nu_n(R)$ decreases monotonically as R increases, and it can be shown that it approaches the n-th eigenvalue λ_n for the infinite domain. Every positive number is actually the limit point of eigenvalues $\nu_n(R)$; for, in the case of a finite domain there are arbitrarily large positive eigenvalues $\nu_n(R)$, and if we let n increase with R in a suitable manner we can approximate any positive number.

We can prove that the eigenvalues *have a limit point at zero*; the method is similar to that used to show that the eigenvalues for a finite domain become infinite. No explicit knowledge of the solutions of a special problem is needed.

If the eigenvalues have a fixed negative upper bound, we can construct a sequence of functions $\varphi_1, \varphi_2, \cdots, \varphi_\nu, \cdots$ for which (a) the integrals $D[\varphi] = \int (\varphi_x^2 + \varphi_y^2 + \varphi_z^2)\, dg$ and $H[\varphi] = \int \varphi^2\, dg$ remain less than a fixed upper bound, (b) the integral $F[\varphi] = \int V\varphi^2\, dg$ always remains greater than a fixed positive bound, and (c) the orthogonality relations $F[\varphi_\nu, \varphi_\mu] = 0$ $(\nu \neq \mu)$ are satisfied. Using a lemma (to be

proved below), we see from property (a) that a subsequence φ_{ν_n} can be chosen from the functions φ_ν with the property $F[\varphi_{\nu_n} - \varphi_{\nu_m}] \to 0$ as $n, m \to \infty$. But since $F[\varphi_{\nu_n}, \varphi_{\nu_m}] = 0$, this would imply the relation $F[\varphi_{\nu_n}] + F[\varphi_{\nu_m}] \to 0$, which contradicts property (b).

The sequence of functions φ_ν is constructed as follows: We begin with the above variational problem (43), which gives the first eigenvalue λ_1. We can find a function φ_1 for which the two relations

$$J[\varphi_1] = D[\varphi_1] - F[\varphi_1] \leq \lambda_1 + \epsilon \qquad (\epsilon > 0)$$

and

$$H[\varphi_1] = 1$$

hold. We now turn to the variational problem (43), (44), which furnishes the second eigenvalue λ_2; if we impose the auxiliary condition

$$\int V\varphi\varphi_1 \, dg = F[\varphi, \varphi_1] = 0,$$

we obtain as the minimum (by the maximum-minimum property) a value which is certainly no greater than λ_2. Thus it is possible to find a function φ_2 such that

$$D[\varphi_2] - F[\varphi_2] \leq \lambda_2 + \epsilon$$

while at the same time

$$H[\varphi_2] = 1, \qquad F[\varphi_1, \varphi_2] = 0.$$

Continuing in this way we obtain a sequence of functions $\varphi_1, \varphi_2, \cdots \varphi_\nu, \cdots$ for which

$$D[\varphi_\nu] - F[\varphi_\nu] \leq \lambda_\nu + \epsilon, \qquad H[\varphi_\nu] = 1, \qquad F[\varphi_\mu, \varphi_\nu] = 0 \qquad (\mu = 1, 2, \cdots, \nu - 1).$$

Now if all the numbers λ_ν were less than the bound -2ϵ, then all these functions would satisfy the inequality

$$D[\varphi_\nu] - F[\varphi_\nu] \leq -\epsilon. \tag{47}$$

From this inequality it follows, in the first place, that $D[\varphi_\nu]$ remains bounded; for, by (45), (46) we have

$$F[\varphi] \leq 4aD[\varphi] + bH[\varphi]$$

and therefore

$$(1-4a)D[\varphi_\nu] \leq b.$$

On the other hand, (47) implies that $F[\varphi_\nu] \geq \epsilon$. This proves that the functions φ_ν actually have the desired properties.

It remains to prove the lemma mentioned above: Given a sequence of functions φ_ν for which $D[\varphi]$ and $H[\varphi]$ are bounded, we can find a subsequence φ_{ν_n} such that the relation

$$F[\varphi_{\nu_n} - \varphi_{\nu_m}] \to 0 \qquad (n, m \to \infty)$$

holds.

This theorem is a generalization of the lemma of Rellich mentioned earlier (§2, 2) which made it possible to prove that the eigenvalues for a finite domain become infinite. We now restrict ourselves to the case in which the function V is regular at the origin. (If V is singular at the origin of degree less than two, then similar estimates make it possible to obtain the desired result.)

To prove the lemma, we exclude infinity by means of a sequence of spheres K_i of radii R_i. In virtue of the earlier lemma (page 414) we can select a subsequence φ_{1n} from the functions φ_ν for which $F[\varphi_{1n} - \varphi_{1m}]$ approaches zero provided that the integral is taken only over the interior of the sphere K_1. From this sequence φ_{1n} we can again select a subsequence φ_{2n} for which the integral $F[\varphi_{2n} - \varphi_{2m}]$ taken over the sphere K_2 approaches zero. We continue in this manner obtaining sequences φ_{in} for each i, and form the diagonal sequence φ_{nn}, which we shall denote by φ_{ν_n}. We know that for this sequence the integral $F[\varphi_{\nu_n} - \varphi_{\nu_m}]$, taken over any one of the spheres K_i, approaches zero. In order to show that the same is true when the integral is extended over the whole space, we have only to show that the integral over the exterior of the sphere K_i is always less than some bound independent of n and m which approaches zero as R increases without bound. To do this, we remark that for sufficiently large R and for $r \geq R$ we have assumed (42): $V \leq B/r^\beta \leq B/R^\beta$, and that therefore the integral taken over the exterior of the sphere of radius R satisfies the inequality

$$F[\varphi_{\nu_n} - \varphi_{\nu_m}] \leq \frac{B}{R^\beta} H[\varphi_{\nu_n} - \varphi_{\nu_m}] \leq \frac{4B}{R^\beta}.$$

This proves our assertion.

§6. *Nodes of Eigenfunctions*

It was possible to make precise statements with regard to the general behavior of the eigenvalues; however the investigation of the general properties of the eigenfunctions offers greater difficulties. This is not surprising, in view of the large number of classes of functions defined by means of eigenvalue problems. A few special cases will be studied in the following chapter, while in the present section we shall be concerned with a more general investigation of eigenfunctions.

The nodes, i.e. those points of the fundamental domain G at which some eigenfunction vanishes, are of particular interest (cf. Ch. V, §5). In dealing with problems in one, two, three, etc. dimensions, we speak of *nodal points*, *nodal curves*, *nodal surfaces*, respectively; in general we use the term *nodes*.[1]

We remark that the first eigenfunction of an eigenvalue problem can have no nodes in the interior of the fundamental domain (the proof of this will follow directly from the theorem given below). It must therefore have the same sign everywhere, and hence *every other eigenfunction* orthogonal to it *must have nodes*.

It is possible to make a number of general statements concerning the position and density of the nodes. For example, consider the differential equation $\Delta u + \lambda u = 0$ with the boundary condition $u = 0$. If G' is a domain which lies entirely in G and contains no nodal points of u_n, we consider the smallest subdomain G'' of G containing G' and bounded by nodes of the function u_n. For this domain G'' the function u_n must be the first eigenfunction, λ_n the smallest eigenvalue. On the other hand, according to Theorem 3 the first eigenvalue of G'' cannot be greater than the first eigenvalue γ of G', and thus $\gamma \geq \lambda_n$. For example, if G' is a circle of radius a then $\gamma = \tau^2$, where τ is the smallest root of the equation $J_0(a\tau) = 0$. We thus have $\gamma = k_{0,1}^2/a^2$ where, as in Ch. V, §5, 5, we denote by $k_{0,1}$ the first zero of the zero-th Bessel function. We therefore obtain $a^2 \leq k_{0,1}^2/\lambda_n$; this relation tells us as much about the density of the net of nodal lines as can in general be expected. If we recall the asymptotic relation $\lambda_n \sim 4\pi n/f$ of §4, we see that if n is sufficiently large every circle whose area is greater than $k_{0,1}^2 f/4n$ must contain

[1] We postulate that for the differential equation under consideration the nodes are piecewise smooth curves or surfaces and decompose the fundamental domain into subdomains with piecewise smooth boundaries.

nodal lines of the n-th eigenfunction. If instead of a circle we consider a square of side a, we find correspondingly $a^2 \leq 2\pi^2/\lambda_n$. The reader will be able to derive entirely analogous statements for other problems in one or several variables.

In addition, it is possible to prove the following general theorem concerning the nodes of an eigenfunction: *Given the self-adjoint second order differential equation* $L[u] + \lambda\rho u = 0$ $(\rho > 0)$ *for a domain* G *with arbitrary homogeneous boundary conditions; if its eigenfunctions are ordered according to increasing eigenvalues, then the nodes of the* n*-th eigenfunction* u_n *divide the domain into no more than* n *subdomains. No assumptions are made about the number of independent variables.*[1]

For simplicity we suppose G is a domain in the x, y-plane and the boundary condition is $u = 0$. Let λ_n be the n-th eigenvalue, i.e. the maximum-minimum of the associated integral $D[\varphi]$ under the prescribed boundary condition and the auxiliary conditions

$$\iint_G \rho\varphi^2 \, dx \, dy = 1, \tag{48}$$

$$\iint_G \rho\varphi v_i \, dx \, dy = 0 \quad (i = 1, 2, \cdots, n-1). \tag{49}$$

We suppose that the nodes of the associated eigenfunction u_n decompose the domain G into more than n subdomains $G_1, G_2, \cdots, G_{n+1}, \cdots$, and define n functions $w_1, w_2, \cdots, w_n$, such that w_i coincides up to a normalizing factor with u_n in the subdomain G_i and vanishes outside G_i, and such that

$$\iint_G \rho w_i^2 \, dx \, dy = 1.$$

If we form a linear combination $\varphi = c_1 w_1 + c_2 w_2 + \cdots + c_n w_n$, which itself satisfies the normalizing condition

$$\iint_G \rho\varphi^2 \, dx \, dy = c_1^2 + c_2^2 + \cdots + c_n^2 = 1,$$

we see at once, integrating by parts, that it satisfies the equation

$$D[\varphi] = \lambda_n,$$

[1] See Courant, Ein allgemeiner Satz zur Theorie der Eigenfunktionen selbstadjungierter Differentialausdrücke.

since w_i satisfies $L[w_i] + \lambda_n \rho w_i = 0$. Now since for any given functions v_i the coefficients c_i can be determined in such a way that φ fulfills conditions (49) in addition to (48), the n-th eigenvalue λ_n' for the domain $G' = G_1 + G_2 + \cdots + G_n$ of the same differential equation with the boundary condition $u = 0$ can be no larger than λ_n; it is precisely equal to λ_n, since by Theorem 2, §2, 1, it also cannot be less than λ_n. From this it follows by Theorem 3 that for every subdomain G'' of G which contains G' the n-th eigenvalue is precisely equal to λ_n. Let $u_n^{(1)}, u_n^{(2)}, \cdots, u_n^{(m)}$ be the eigenfunctions obtained in this way, for an arbitrary number m of such domains $G', G'', G''', \cdots, G^{(m)}$ each containing the preceding domain. If these eigenfunctions are continued into the exterior of the corresponding subdomain of G by the requirement that they vanish there identically, then they form a system of m linearly independent[1] functions which are all solutions in G of the differential equation $L[u_n^{(i)}] + \lambda_n \rho u_n^{(i)} = 0$. We can determine a linear combination

$$\varphi = \gamma_1 u_n^{(1)} + \cdots + \gamma_m u_n^{(m)}$$

with coefficients γ_i, which do not vanish everywhere, in such a way that the $m - 1$ conditions

$$\iint_G \rho \varphi v_i \, dx \, dy = 0 \quad (i = 1, 2, \cdots, m - 1)$$

are satisfied; since φ cannot vanish identically because of the linear independence of the $u_n^{(i)}$, we can normalize it according to (48) by multiplying by a suitable factor. But then, because of the maximum-minimum property of the m-th eigenfunction, we must have

$$D[\varphi] \geq \lambda_m .$$

On the other hand, integrating by parts, we find

$$D[\varphi] = \lambda_n .$$

However, since $\lim_{k \to \infty} \lambda_k = \infty$ we have $\lambda_m > \lambda_n$, for sufficiently large m. We thus obtain a contradiction, which demonstrates the

[1] It is immediately obvious that these functions are linearly independent if one remembers that $u_n^{(h)}$ cannot be identically zero in any subdomain of $G^{(h)}$. This fact, which in the case of ordinary differential equations follows from the uniqueness theorem, is for partial differential equations a consequence of their elliptic character, to which we shall return in Volume II.

impossibility of the above assumption of more than n domains G_1, G_2, $\cdots$. It need hardly be said that this proof is valid for any number of variables.[1]

For the special case of the Sturm-Liouville eigenvalue problem $(py')' - qy + \lambda\rho y = 0$, the general theorem just proved may be strengthened in a remarkable way. Here we can show in addition that the n-th eigenfunction divides the fundamental domain into no less than n parts; thus we have the theorem: *The n-th eigenfunction for a Sturm-Liouville problem divides the fundamental domain into precisely n parts by means of its nodal points.* This is usually proved by continuity considerations which we now sketch.

For simplicity we restrict ourselves to the differential equation $y'' + \lambda\rho y = 0$. By $y(x, \lambda)$ we denote a solution of this equation which depends continuously on the parameter λ and vanishes at $x = 0$; we obtain the identity

$$y(x, \lambda_1)y'(x, \lambda) - y(x, \lambda)y'(x, \lambda_1) = (\lambda_1 - \lambda) \int_0^x \rho y(x, \lambda)y(x, \lambda_1)\, dx.$$

If $x = \xi$ is a positive zero of $y(x, \lambda)$, it follows that

$$y(\xi, \lambda_1)y'(\xi, \lambda) = (\lambda_1 - \lambda) \int_0^\xi \rho y(x, \lambda)y(x, \lambda_1)\, dx.$$

Now let λ_1 be greater than λ and so close to λ that the integral on the right remains positive. Then $y(\xi, \lambda_1)$ and $y'(\xi, \lambda)$ must have the same sign. If we assume that at $x = \xi$ the function $y(x, \lambda)$ changes from negative to positive values and thus that $y'(\xi, \lambda)$ is positive—$y'(\xi, \lambda)$ and $y(\xi, \lambda)$ cannot both vanish at the same time—then $y(\xi, \lambda_1)$ is also positive. Since $y(x, \lambda_1)$ differs arbitrarily little from $y(x, \lambda)$ for sufficiently small $\lambda_1 - \lambda$ and must, therefore, pass from negative to positive values in the neighborhood of $x = \xi$, a zero of $y(x, \lambda_1)$ lies at the left[2] of ξ and we can state: *As λ increases continuously, all the zeros of the function $y(x, \lambda)$ decrease.* The first eigenfunction has zeros only at the two ends of the fundamental domain. As λ passes from the first eigenvalue to the second eigenvalue, the second zero moves from the right into the interior of the interval, continuing until the end

[1] The theorem just proved may be generalized as follows: Any linear combination of the first n eigenfunctions divides the domain, by means of its nodes, into no more than n subdomains. See the Göttingen dissertation of H. Herrmann, Beiträge zur Theorie der Eigenwerte und Eigenfunktionen, 1932.

[2] The fact that y and y' can never vanish at the same point implies that zeros neither appear nor disappear between 0 and ξ as λ increases.

point of the interval becomes a third zero of the function, etc., and thus the theorem is evident.[1]

We were able to prove this fact because we were dealing with an ordinary differential equation. In the case of eigenvalue problems of partial differential equations, arbitrarily large values of n may exist for which the nodes of the eigenfunctions u_n subdivide the entire fundamental domain into only two subdomains. Simple examples of this[2] are offered by the equation $\Delta u + \lambda u = 0$ for a square: $0 \leq x \leq \pi$, $0 \leq y \leq \pi$. In this case, it is easily seen that the eigenfunctions $\sin 2rx \sin y + \mu \sin x \sin 2ry$ belonging to the eigenvalues $\lambda = 4r^2 + 1$ (where μ is a positive constant sufficiently close to 1) possess only a single nodal curve. Figures 7 and 8 show how these nodal curves arise from slight variation of a system of curves for the case $r = 12$.

§7. *Supplementary Remarks and Problems*

1. Minimizing Properties of Eigenvalues. Derivation from Completeness. The completeness of the eigenfunctions obtained in our variational problems has been used to prove that they constitute the

[1] We can avoid continuity methods by basing the proof on the following theorem, which is not limited to one independent variable: Let u be a solution of $L[u] + \lambda\rho u = 0$ which is twice continuously differentiable in a closed region B. If u vanishes on the boundary Γ of B without changing sign in the interior, and if v is a solution of $L[v] + \mu\rho v = 0$ where $\mu > \lambda$, then v must change sign in B. (Naturally we exclude the case in which u or v vanishes identically in B.) We prove this at once by concluding with the aid of Green's formula—say for the case of two independent variables—that

$$\iint_B (vL[u] - uL[v])\, dx\, dy = (\mu - \lambda) \iint_B \rho uv\, dx\, dy = \int_\Gamma v \frac{\partial u}{\partial n}\, ds,$$

where $\partial/\partial n$ denotes differentiation with respect to the outer normal. Without loss of generality we may assume that u and v take positive values in B; since we have $\partial u/\partial n \leq 0$ on Γ, the expression on the right in the above equation is not positive, while the middle expression would have to be positive if v did not change sign in B.

If we apply this result to the Sturm-Liouville problem with vanishing boundary values, we see that of two eigenfunctions, the one having the larger number of zeros must belong to the larger eigenvalue. For, an interval between two suitably chosen zeros of the eigenfunction with fewer zeros must contain as a proper subinterval such an interval determined by two zeros of the other eigenfunction. Since the first eigenfunction has no zeros in the interior, the n-th (which can have no more than $n - 1$) must accordingly have $n - 1$ zeros in the fundamental domain.

[2] See A. Stern, Bemerkungen über asymptotisches Verhalten von Eigenwerten und Eigenfunktionen, Dissertation, Göttingen, 1925.

totality of solutions of the corresponding differential equations. Conversely, we may (e.g. in the case of the trigonometric and Legendre functions) start with a complete system of functions which are solutions of the eigenvalue problem for a differential equation. We shall show that such systems of functions coincide with the systems defined by means of the extremal properties: Suppose we are dealing with the differential equation

$$L[u] + \lambda\rho u = 0$$

for the two-dimensional domain G and the boundary condition $u = 0$. Let the eigenfunctions of the differential equation problem be $u_1, u_2, \cdots$ and the associated eigenvalues $\lambda_1, \lambda_2, \cdots$. Consider

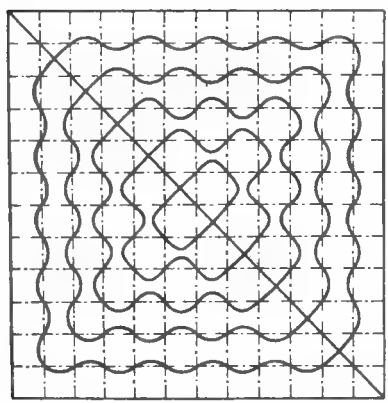


Figure 7

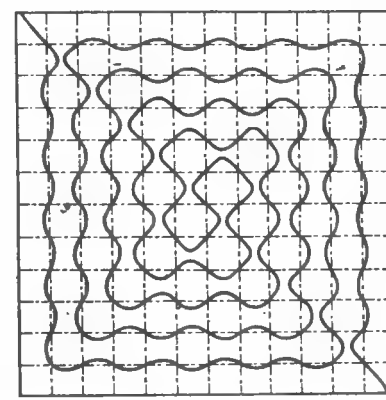


Figure 8

any function φ with continuous first derivatives and piecewise continuous second derivatives in G which vanishes on the boundary Γ and satisfies the conditions

$$\iint_G \rho\varphi^2\, dx\, dy = 1, \tag{50}$$

$$\iint_G \rho\varphi u_i\, dx\, dy = 0 \quad (i = 1, 2, \cdots, n-1). \tag{51}$$

We now show that φ satisfies the inequality

$$D[\varphi] \geq \lambda_n .$$

For, because of the boundary condition $\varphi = 0$, Green's formula yields

$$D[\varphi] = -\iint_G \varphi L[\varphi]\, dx\, dy;$$

in addition, the completeness relation (cf. formula (23a), page 426). applied to the functions φ and $L[\varphi]/\rho$, leads to the relation

$$D[\varphi] = -\sum_{i=1}^{\infty} \gamma_i \iint_G u_i L[\varphi]\, dx\, dy \tag{52}$$

where

$$\gamma_i = \iint_G \rho \varphi u_i\, dx\, dy.$$

It follows from (52), with the aid of Green's formula and the relation $L[u_i] = -\lambda_i \rho u_i$, that

$$D[\varphi] = \sum_{i=1}^{\infty} \lambda_i \gamma_i^2. \tag{53}$$

Now, since (51) implies

$$\gamma_i = 0 \qquad \text{for } i = 1, 2, \cdots, n-1,$$

and since from (50), together with the completeness relation, we have

$$\sum_{i=1}^{\infty} \gamma_i^2 = 1,$$

it follows immediately that if the λ_i are ordered according to increasing magnitude, we have

$$D[\varphi] \geq \lambda_n .$$

Moreover, as shown earlier, by a simple calculation we obtain

$$D[u_n] = \lambda_n ,$$

which means we have proved the minimizing property of the n-th eigenfunction with respect to the function φ. The same considerations may be applied if we assume only that the functions φ are continuous and have piecewise continuous first derivatives; for, a function of this kind, together with its derivative, may always be approximated by a function of the class defined above in such a way that the cor-

responding integrals $D[\varphi]$ differ by an arbitrarily small amount (in this connection, see the remarks in Ch. IV, §3, 7).

2. Characterization of the First Eigenfunction by Absence of Nodes. The first eigenfunction has been characterized by the property that it does not vanish. We shall now use an important method introduced by Jacobi to investigate this property (method of multiplicative variation).

We limit ourselves to the equation

$$\Delta u - qu + \lambda u = 0.$$

We wish to prove: If there exists a solution u of this equation which vanishes on the boundary Γ of a domain G but nowhere in the interior, then

$$\mathfrak{D}[\varphi] = \iint_G (\varphi_x^2 + \varphi_y^2 + q\varphi^2)\, dx\, dy \geq \lambda \iint_G \varphi^2\, dx\, dy$$

for all admissible functions φ, where the equality holds only for $\varphi = \text{const} \times u$. We suppose every such function φ to be represented in the form

$$\varphi = u\eta,$$

which is possible, since u does not vanish in G; we have

$$\mathfrak{D}[\varphi] = \iint_G [u^2(\eta_x^2 + \eta_y^2) + 2uu_x\eta\eta_x + 2uu_y\eta\eta_y + (u_x^2 + u_y^2)\eta^2 + qu^2\eta^2]\, dx\, dy.$$

By observing $2\eta\eta_x = (\eta^2)_x$, $2\eta\eta_y = (\eta^2)_y$ and integrating by parts, we obtain

$$\mathfrak{D}[\varphi] = \iint_G [u^2(\eta_x^2 + \eta_y^2) - u\Delta u\,\eta^2 + qu^2\eta^2]\, dx\, dy$$

since all the integrals over the boundary vanish. Making use of the differential equation for u, we find

$$\mathfrak{D}[\varphi] = \iint_G [u^2(\eta_x^2 + \eta_y^2) + \lambda u^2\eta^2]\, dx\, dy \geq \lambda \iint_G u^2\eta^2\, dx\, dy = \lambda \iint_G \varphi^2\, dx\, dy,$$

where the equality holds only for $\eta_x = \eta_y = 0$, i.e. for $\eta = \text{const.}$, q.e.d.

3. Further Minimizing Properties of Eigenvalues. The reader may prove the following theorem: The problem of minimizing the integral expression

$$\mathsf{D}[v_1, v_2, \cdots, v_n] = \mathfrak{T}[v_1] + \mathfrak{T}[v_2] + \cdots + \mathfrak{T}[v_n],$$

in which we admit all systems of n mutually orthogonal normalized functions with piecewise continuous derivatives in the fundamental domain G, is solved by the functions $v_i = u_i$ or by any system of functions which arise from these functions through an orthogonal transformation. Here the functions $u_1, u_2, \cdots, u_n$ stand for the first n eigenfunctions for the domain. The minimum of $\mathsf{D}[v_1, v_2, \cdots, v_n]$ is equal to the sum of the first n eigenvalues, $\lambda_1 + \lambda_2 + \cdots + \lambda_n$.

The following theorem can also be proved: Let $v_1, v_2, \cdots, v_{n-1}$ be continuous functions in G, and let $d\{v_1, v_2, \cdots, v_{n-1}\}$ be the lower bound of the integral expression $\mathfrak{T}[\varphi]$ where, in addition to satisfying the usual continuity conditions, φ is subject to the single auxiliary condition

$$\iint_G \rho\varphi^2 \, dx \, dy - \sum_{i=1}^{n-1} \left(\iint_G \rho\varphi v_i \, dx \, dy \right)^2 = 1.$$

Then the n-th eigenvalue λ_n equals the maximum of $d\{v_1, v_2, \cdots, v_{n-1}\}$ and is assumed for $v_1 = u_1, v_2 = u_2, \cdots, v_{n-1} = u_{n-1}; \varphi = u_n$.

This formulation is interesting since it uses only one quadratic auxiliary condition and does not require linear conditions; however, the auxiliary condition has a form which is unusual for isoperimetric problems. It will be left as a problem for the reader to carry over this formulation to the corresponding elementary problem for quadratic forms.

Other formulations of the eigenvalue problems, useful in many applications, will be given in connection with the differential equation $\Delta u + \lambda u = 0$ with the boundary condition $u = 0$:

$$H[\varphi] = \iint_G \varphi^2 \, dx \, dy = \text{min. max.}$$

subject to the auxiliary conditions

$$D[\varphi] = \iint_G (\varphi_x^2 + \varphi_y^2) \, dx \, dy = 1,$$
$$D[\varphi, v_i] = 0 \qquad (i = 1, 2, \cdots, n-1),$$

from which one sees immediately the significance of the formulation as a maximum-minimum problem.

Another equivalent problem is that of making the expression

$$\iint_G (\Delta\varphi)^2 \, dx \, dy$$

a maximum-minimum under the same auxiliary conditions, where we again require the functions φ to have continuous first and piecewise continuous second derivatives.

4. Asymptotic Distribution of Eigenvalues. (a) For the differential equation $\Delta\Delta u - \lambda u = 0$ of the vibrating plate, with the boundary conditions $u = 0$ and $\partial u/\partial n = 0$ (clamped plate), we have the asymptotic relation

$$A(\lambda) \sim \frac{f}{4\pi} \sqrt{\lambda},$$

from which

$$\lambda_n \sim \left(\frac{4\pi n}{f}\right)^2$$

follows. Here, as earlier, $A(\lambda)$ is the number of eigenvalues less than the bound λ, λ_n denotes the n-th eigenvalue, and f the area of the plate. We may state, therefore: The n-th eigenvalue of the clamped plate, as n increases, is asymptotically equal to the square of the n-th eigenvalue of the clamped membrane. In particular, it depends only on the size and not on the shape of the plate. An analogous statement holds in three dimensions.[1]

(b) Derive the laws of asymptotic distribution of eigenvalues for the Sturm-Liouville equation (see the results of §2, 3) and also for ordinary differential equations of fourth order, using the methods of §4, 3.

(c) Derive the laws of the asymptotic distribution of eigenvalues for the case of elliptic self-adjoint differential equations arising from an arbitrary definite quadratic variational problem.

5. Parameter Eigenvalue Problems. Carry out the solution of two-parameter eigenvalue problems (see the Lamé problem in Ch. V, §9, 3) using methods of the calculus of variations.

[1] See Courant, Über die Schwingungen eingespannter Platten.

6. Boundary Conditions containing Parameters. Eigenvalue problems with the parameter in the boundary condition (cf. Ch. V, §16, 4) may also be solved easily with the aid of methods of the calculus of variations. For the differential equation $\Delta u = 0$ and the boundary condition $\partial u/\partial n = \lambda u$, we have to minimize an integral of the form

$$\iint_G (\varphi_x^2 + \varphi_y^2)\, dx\, dy$$

where the integral of φ^2 over the boundary satisfies the condition

$$\int_\Gamma \varphi^2\, ds = 1$$

and suitable linear auxiliary conditions are imposed. Further development of this idea is left to the reader.

If G is the unit circle, the solutions of this problem are given by the potential functions $r^n \cos n\theta$, $r^n \sin n\theta$; the eigenvalues are $\lambda_n = n$.

In the general case, it is easily seen with the aid of the methods developed in the present chapter that λ_n is of the order of n. It therefore follows from §3, 1 that the eigenfunctions are complete with respect to the expression $\mathfrak{H}[\varphi] = \int_\Gamma \varphi^2\, ds$. In other words, the boundary values of the eigenfunctions considered as functions of s form a complete system from which, in turn, it may be concluded that every potential function which is regular in G may be approximated in the mean by its eigenfunctions.

7. Eigenvalue Problems for Closed Surfaces. The eigenvalue problem of the Laplace spherical harmonics is an example of a problem on a closed surface, for which the condition of regularity over the entire surface replaces boundary conditions. The methods of the present chapter show that these eigenvalue problems are closely related to minimum or to maximum-minimum problems for a quotient $\mathfrak{D}/\mathfrak{H}$; $\mathfrak{D}$ is a quadratic expression formed with the derivatives of φ and $\mathfrak{H}[\varphi]$ is a positive definite quadratic expression which does not contain derivatives and has the closed surface as the domain of integration. This theory may be carried over to other quadratic differential expressions on closed surfaces.

8. Estimates of Eigenvalues when Singular Points Occur. In §2, 4 we treated singular points in connection with Bessel eigenvalue

problems; Bessel functions of zero-th order required special treatment using particular properties of the Bessel functions. This will no longer be necessary because we shall now develop a more general method.

We consider the eigenvalue problem associated with the expressions

$$D[\varphi] = \int_0^1 x\varphi'^2\,dx, \qquad H[\varphi] = \int_0^1 x\varphi^2\,dx$$

with no boundary conditions at $x = 0$, and with the boundary condition $\varphi(1) = 0$ at $x = 1$. After introducing $\sqrt{x}\,\varphi$ as new dependent variable, we find without difficulty the estimate $\lambda_n \leq n^2\pi^2$ for the n-th eigenvalue λ_n of our problem. Thus for the number $A(\lambda)$ of eigenvalues less than λ we have $A(\lambda) \geq \sqrt{\lambda}/\pi$.

In order to estimate λ_n from below, and thus find an upper bound for $A(\lambda)$—this is our specific objective here—we choose an arbitrarily small positive number ϵ between 0 and 1 and note that $A(\lambda) \leq B_1(\lambda) + B_2(\lambda)$, where B_1 and B_2 denote the number of eigenvalues less than λ of the expressions

$$D_1 = \int_0^\epsilon x\varphi'^2\,dx, \qquad H_1 = \int_0^\epsilon x\varphi^2\,dx$$

and

$$D_2 = \int_\epsilon^1 x\varphi'^2\,dx, \qquad H_2 = \int_\epsilon^1 x\varphi^2\,dx,$$

respectively. Here we no longer require that the function φ be continuous at the point $x = \epsilon$, so that in both cases the point $x = \epsilon$ occurs as a free end point. Using the methods developed in this chapter we find the asymptotic relation $B_2(\lambda)/\sqrt{\lambda} \to (1 - \epsilon)/\pi$ for $B_2(\lambda)$; it remains to estimate $B_1(\lambda)$. We majorize H_1 by H_1^* and minorize D_1 by D_1^*:

$$H_1^* = \epsilon\int_0^\epsilon \varphi^2\,dx \qquad D_1^* = \int_0^\epsilon x\left(1 - \frac{x}{\epsilon}\right)\varphi'^2\,dx.$$

Using an obvious notation, we have $B_1(\lambda) < B_1^*(\lambda)$. On the other hand, we can write down the eigenfunctions and eigenvalues of this new eigenvalue problem explicitly. To do this, we map the interval $0 \leq x \leq \epsilon$ onto the interval $-1 \leq \xi \leq 1$ by the transformation

$x = (1 + \xi)\epsilon/2$. For the eigenfunctions we obtain the Legendre polynomials in ξ, and for the eigenvalues the numbers $n(n + 1)/\epsilon^2$. Hence $B_1(\lambda) \leq B_1^*(\lambda) \leq \epsilon(1 + \delta)\sqrt{\lambda}$, where δ approaches zero with increasing λ. Summing up our results, we obtain almost immediately the asymptotic relation

$$\lim_{\lambda \to \infty} \frac{A(\lambda)}{\sqrt{\lambda}} = \frac{1}{\pi}$$

since ϵ could be chosen arbitrarily small.

9. Minimum Theorems for the Membrane and Plate. Of all clamped membranes or plates with given perimeter or area and given constant density and elasticity, those with the circle as boundary have the lowest fundamental tone. (For proofs, see the first paper cited below for the case of a given constant perimeter, and the works of G. Faber[1] and E. Krahn[2] for the case of a given constant area.)

10. Minimum Problems for Variable Mass Distribution. The reader may prove the following theorems:

The fundamental tone of a stretched string of given uniform tension along which a given mass has been distributed is lowest when the entire mass is concentrated at the midpoint.

Prove the analogous results for the membrane and plate.

11. Nodal Points for the Sturm-Liouville Problem. Maximum-minimum Principle. The theorem of §6, which states that the n-th eigenfunction divides the fundamental domain into n parts by means of its zeros, can also be proved on the basis of the following considerations.[3] Given a string capable of vibrating, if $n - 1$ arbitrarily chosen inner points are fixed, the fundamental tone of the resulting system consisting of n independent strings is identical with the lowest of the fundamental tones of the subsystems (see §1, 3). The fundamental vibration of the decomposed system is then the fundamental vibration of the subsystem in question while the remaining subsystems are constrained to the rest position. If we vary the prescribed nodal points, the fundamental tone of the decomposed

[1] G. Faber, Beweis, dass unter allen homogenen Membranen von gleicher Fläche und gleicher Spannung die kreisförmige den tiefsten Grundton gibt, S.-Ber. Bayr. Akad. Wiss. (math.-phys. Kl.), 1923, pp. 169–172.

[2] E. Krahn, Über eine von Rayleigh formulierte Minimaleigenschaft des Kreises, Math. Ann., Vol. 94, 1925, pp. 97–100.

[3] See K. Hohenemser, Praktische Wege zur angenäherten Schwingungsberechnung elastischer Systeme, Ingenieur-Archiv, Vol. 1, No. 3, 1930, pp. 1–24.

system reaches a maximum when the n subsystems all have the same fundamental tone. For, if two neighboring subsystems had different fundamental tones, we could raise the fundamental tone of one and lower that of the other by moving the nodal point which they have in common so that both subsystems would have equally high fundamental tones. Now, in the extremal case considered, the fundamental vibration of the decomposed system may be represented by a continuously differentiable function which represents an eigenfunction of the original unrestrained system associated with the vibration frequency in question and vanishing at the $n - 1$ points corresponding to the maximum fundamental tone. Hence: *If a string is fixed at $n - 1$ points, and we wish to choose these points in such a way that the fundamental tone of the resulting system is as high as possible, we find that the solution is given by an eigenfunction of the original system having $n - 1$ zeros at inner points.* If we call the eigenvalues thus obtained μ_n and the associated eigenfunctions v_n, then $\mu_{n+1} \geq \mu_n$ since there exists an interval, determined by two suitable neighboring zeros of v_n, which contains the interval between two zeros of v_{n+1} as a proper subinterval, and since a reduction in the length of the interval produces a rise in the fundamental tone. (See page 455, footnote).

If, as before, we denote the eigenvalues of the string, ordered with respect to increasing magnitude, by λ_n, we see that we always have $\mu_n \geq \lambda_n$ since the μ_n must certainly be contained among the λ_n. On the other hand, the restriction of a prescribed nodal point is only a special or limiting case of a linear auxiliary condition such as we have considered in §1, 4 for the variational problems defining the eigenvalues λ_n. If we limit ourselves to such special auxiliary conditions the maximum-minimum, i.e. the number μ_n, cannot be greater than the maximum of the minimum when arbitrary linear auxiliary conditions are admitted, i.e. not greater than λ_n. Hence $\mu_n \leq \lambda_n$, and therefore, in view of the above result, we have $\mu_n = \lambda_n$. This completes the proof of the theorem on the zeros of Sturm-Liouville eigenfunctions.

References

Courant, R., Beweis des Satzes, dass von allen homogenen Membranen gegebenem Umfanges und gegebener Spannung die kreisförmige den tiefsten Grundton besitzt. Math. Zeitschr., Vol. 1, 1918, pp. 321–328.

——, Über die Eigenwerte bei den Differentialgleichungen der mathematischen Physik. Ibid., Vol. 7, 1920, pp. 1–57.

——, Über die Schwingungen eigenspannter Platten. Ibid., Vol. 15, 1922, pp. 195–200.

——, Ein allgemeiner Satz zur Theorie der Eigenfunktionen selbstadjungierter Differentialausdrücke. Nachr. Ges. Göttingen (math.-phys. Kl.), 1923, Session of July 13.

——, Über die Anwendung der Variationsrechnung in der Theorie der Eigenschwingungen und über neue Klassen von Funktionalgleichungen. Acta Math., Vol. 49, 1926, pp. 1–68.

Kneser, A., Die Integralgleichungen und ihre Anwendungen in der mathematischen Physik. 2nd ed., F. Vieweg und Sohn, Braunschweig, 1922.

Liouville, J., Mémoire sur le développement des fonctions ou parties de fonctions en séries dont les divers termes sont assujetis à satisfaire à une même équation différentielle du second ordre contenant un paramètre variable. J. de math. pures et appl., Ser. 1, Vol. 1, 1836, pp. 253–265. Ibid., Vol. 2, 1837, pp. 16–35, 418–436.

Weyl, H., Das asymptotische Verteilungsgesetz der Eigenwerte linearer partieller Differentialgleichungen (mit einer Anwendung auf die Theorie der Hohlraumstrahlung). Math. Ann., Vol. 71, 1912, pp. 441–479.

——, Über die Abhängigkeit der Eigenschwingungen einer Membran von deren Begrenzung. Journ. f. d. reine u. angew. Math., Vol. 141, 1912, pp. 1–11.

——, Über das Spektrum der Hohlraumstrahlung. Ibid., pp. 163–181.

——, Über die Randwertaufgabe der Strahlungstheorie und asymptotische Spektralgesetze. Ibid., Vol. 143, 1913, pp. 177–202.

Richardson, R. G. D., Das Jacobische Kriterium der Variationsrechnung und die Oszillationseigenschaften linearer Differentialgleichungen zweiter Ordnung. First communication, Math. Ann., Vol. 68, 1910, p. 279. Second communication, Ibid., Vol. 71, 1911–12, p. 214.

——, Über die notwendigen und hinreichenden Bedingungen für das Bestehen eines Kleinschen Oszillationstheorems. Ibid., Vol. 73, 1918, p. 289.

In an unpublished manuscript, submitted in 1914 to the editors of Mathematische Annalen in the form of an abstract, Richardson treated the eigenvalue problem of elliptic differential equations with results similar to some of those of the present chapter. In particular, Richardson's paper dealt with the behavior of the eigenvalues and the coefficients in the boundary condition with respect to the growth of the domain and their independence of the coefficients of the differential equation; he also treated theorems concerning the zeros of the eigenfunctions.

CHAPTER VII

Special Functions Defined by Eigenvalue Problems

§1. *Preliminary Discussion of Linear Second Order Differential Equations*

In the present chapter we shall investigate certain classes of functions which have already been defined, Bessel functions, Legendre functions, and general Laplace spherical harmonics. Our point of view will be somewhat more general than in the preceding chapters; we shall permit the independent variable to assume complex values and use the methods of function theory. Furthermore, we shall consider not only the functions mentioned but also the totality of solutions of the corresponding differential equations. We assume it to be known that: linear differential equations of this type, even in the case of a complex independent variable $z = x + iy$, possess two linearly independent solutions, the general solution is a linear combination of these solutions with constant coefficients, and all solutions are regular analytic functions of z except at certain fixed singular points determined by the coefficients. Many important classes of functions can be defined as solutions of such linear differential equations with analytic coefficients.

To obtain solutions of a linear differential equation

$$L[u] + \mu u = 0$$

in the form of an integral representation, the *method of integral transformation* is often useful; we start with a general outline of this method. In place of the unknown function $u(z)$ we introduce a new unknown function $v(\zeta)$ of the complex variable $\zeta = \xi + i\eta$ by means of equation

$$u(z) = \int_C K(z, \zeta)v(\zeta)\, d\zeta, \tag{1}$$

where the transformation kernel $K(z, \zeta)$ (assumed analytic in each

of the complex variables) and the path of integration C are to be suitably determined. The differential equation takes the form

$$\int_C (L[K] + \mu K)v(\zeta)\, d\zeta = 0;$$

here the differentiation process L refers to the variable z and it is assumed that the process L is interchangeable with integration.

Now we specify K by subjecting $K(z, \zeta)$ to a partial differential equation

$$L[K] = A[K]$$

where $A[K]$ is a linear differential expression containing differentiation only with respect to the variable ζ. Integrating by parts we eliminate these derivatives. The above integral takes the form

$$\int_C K(z, \zeta)(B[v] + \mu v)\, d\zeta;$$

here $B[v]$ denotes the differential expression adjoint to $A[v]$ (see Ch. V, §1). In addition to this integral, a term depending on the boundary occurs, which can be made to vanish by a suitable choice of the path of integration. If the partial differential equation for K, which can be chosen in various ways, as well as the transformed equation

$$B[v] + \mu v = 0$$

can be solved explicitly in such a way that the above assumptions hold, then this method leads to the solution $u(z)$ in the indicated integral form.

In analysis such integral transformations occur in various forms. For example the kernel

$$K(z, \zeta) = e^{z\zeta} \text{ or } e^{iz\zeta}$$

produces the *Laplace transformation*, the kernel

$$K(z, \zeta) = (z - \zeta)^{\alpha}$$

the *Euler transformation* if the path of integration is suitably chosen.

§2. *Bessel Functions*

First we shall discuss Bessel's equation

$$z^2u'' + zu' + z^2u - \lambda^2 u = 0 \tag{2}$$

and all its solutions, considering both z and the parameter λ as complex quantities.

1. Application of the Integral Transformation. We subject equation (2) to the transformation (1). Substitution in the differential equation yields

$$\int_C (z^2K_{zz} + zK_z + z^2K - \lambda^2K)v(\zeta)\,d\zeta = 0.$$

We now require that K satisfy the equation

$$z^2K_{zz} + zK_z + z^2K + K_{\zeta\zeta} = 0,$$

for which the function

$$K(z, \zeta) = e^{\pm iz\sin\zeta}$$

is a single-valued regular solution throughout the z- and ζ-planes. Equation (2) becomes

$$\int_C (K_{\zeta\zeta} + \lambda^2 K)v(\zeta)\,d\zeta = 0$$

or, integrating by parts,

$$\int_C K(z, \zeta)\{v'' + \lambda^2 v\}\,d\zeta + \int_C \frac{\partial}{\partial\zeta}\{K_\zeta v - Kv'\}\,d\zeta = 0.$$

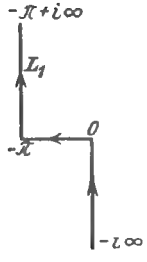


Figure 9

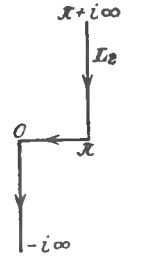


Figure 10

Since the transformed equation $v'' + \lambda^2 v = 0$ has the solutions $e^{\pm i\lambda\zeta}$, we have only to choose the path of integration in a suitable way. We notice that on the vertical portions of the paths denoted by L_1 and L_2 in Figures 9 and 10, the real part of $-iz\sin\zeta$ is negative for $\mathcal{R}_e(z) > 0$ and approaches $-\infty$ exponentially with increasing $|\zeta|$. Therefore, if we set $K(z, \zeta) = e^{-iz\sin\zeta}$, the expression $K_\zeta v - Kv'$ tends to zero in both directions on L_1 and L_2, and we obtain the integrals

$$(3)\qquad H_\lambda^1(z) = -\frac{1}{\pi}\int_{L_1} e^{-iz\sin\zeta + i\lambda\zeta}\,d\zeta,$$

$$H_\lambda^2(z) = -\frac{1}{\pi}\int_{L_2} e^{-iz\sin\zeta + i\lambda\zeta}\,d\zeta$$

as the two solutions of equation (2), the so-called *Hankel functions.* It can be easily verified that these integrals converge for $\mathcal{R}_e(z) > 0$ and satisfy the assumptions necessary for their derivation.

2. Hankel Functions. The Hankel functions $H_\lambda^1(z)$ and $H_\lambda^2(z)$ are defined by the integrals (3) only in the right half-plane $\mathcal{R}_e(z) > 0$. However, it is easy to continue them analytically:

If, for fixed $z = x + iy$, we set

$$f(\zeta) = -iz \sin \zeta + i\lambda\zeta,$$

$$\zeta = \xi + i\eta, \qquad \lambda = a + ib,$$

then

$$\mathcal{R}_e f(\zeta) = y \sin \xi \cosh \eta + x \cos \xi \sinh \eta - b\xi - a\eta,$$

$$\mathcal{I}_m f(\zeta) = -x \sin \xi \cosh \eta + y \cos \xi \sinh \eta + a\xi - b\eta.$$

If, instead of 0 and $-\pi$, we take the abscissas ξ_0 and $-\pi - \xi_0$ for the vertical portions of the path L_1, the integral $\int_{L_1'} e^{f(\zeta)} d\zeta$ taken over the new path L_1' remains convergent for those z for which

$$y \sin \xi_0 - x \cos \xi_0 < 0$$

holds, i.e. for all z in the half-plane bounded by the line

$$y \sin \xi_0 - x \cos \xi_0 = 0.$$

Both paths of integration may be used in that part of this half-plane which also lies in the half-plane $x > 0$. It is clear from Cauchy's integral theorem that they give the same result. In the remaining part, however, the integral over the new path furnishes the analytic continuation of the function $H_\lambda^1(z)$. If, in an appropriate way, we let ξ_0 run through an unbounded sequence of positive and, similarly, through a sequence of negative values, we gradually obtain the complete analytic continuation of the function $H_\lambda^1(z)$; i.e., we obtain a Riemann surface with a branch point at the origin the order of which depends on λ.

For $\xi_0 = -\pi/2$ the horizontal segment of the path of integration vanishes, and for $H_\lambda^1(z)$ we obtain the integral

$$H_\lambda^1(z) = \frac{e^{-i\pi\lambda/2}}{\pi i} \int_{-\infty}^{\infty} e^{iz\cosh\eta - \lambda\eta} \, d\eta,$$

which represents the function in the upper half-plane $\mathcal{I}_m(z) > 0$. If we permit z to become infinite in the sector

$$\delta \leq \arg z \leq \pi - \delta,$$

the integrand tends to zero over the entire path of integration; hence the function $H_\lambda^1(z)$ also tends to zero since the integral converges uniformly in every subset $\mathcal{I}_m(z) \geq \rho > 0$. Similarly, we find the function $H_\lambda^2(z)$ tends to zero as z approaches infinity in the sector

$$-\pi + \delta \leq \arg z \leq -\delta.$$

We therefore have the result:

The Hankel function $H_\lambda^1(z)$ tends to zero when the variable z tends to infinity in a sector $\delta \leq \arg z \leq \pi - \delta$ of the upper half-plane. The Hankel function $H_\lambda^2(z)$ tends to zero when z tends to infinity in a sector $-\pi + \delta \leq \arg z \leq -\delta$ of the lower half-plane.[1]

From the behavior of the Hankel functions at infinity, it follows that: *neither the function $H_\lambda^1(z)$ nor the function $H_\lambda^2(z)$ vanishes identically, and they are linearly independent for every λ.*

To prove this, we show that, as $|z|$ increases, the functions $H_\lambda^2(z)$ and $H_\lambda^1(z)$ increase without limit on the positive and negative imaginary axes, respectively.

We obtain a representation of $H_\lambda^2(z)$ which converges on the positive imaginary axis by taking the values $-\xi_0$ and $\pi + \xi_0$ for the abscissas of the portions of the path L_2', where ξ_0 is any number in the interval $0 < \xi_0 \leq \pi/2$. Since the integrals $\int e^{f(\zeta)} d\zeta$ over these portions converge to zero with increasing y, we need only investigate the remaining integral

$$\int_{\pi+\xi_0}^{-\xi_0} e^{y\sin\xi - b\xi + ia\xi} d\xi;$$

substituting $\xi = \xi' + \pi/2$, this integral becomes

$$\int_0^{(\pi/2)+\xi_0} \cosh b\xi \, e^{y\cos\xi} \cos a\xi \, d\xi.$$

[1] This statement is true only for the initial branches of the functions $H_\lambda^1(z)$ and $H_\lambda^2(z)$; the remaining branches are obtained as linear combinations of the original branches, and do not exhibit the described behavior.

But this integral increases without limit as $y \to \infty$; this is immediately clear if $|a| \leq 1$ and can be seen by somewhat more precise estimates[1] if $|a| > 1$.

The discussion is analogous for the function $H_\lambda^1(z)$ and the negative imaginary axis.

The linear independence of the functions $H_\lambda^1(z)$ and $H_\lambda^2(z)$ implies that the Hankel functions yield the totality of solutions of Bessel's equation. For, every solution may be represented as a linear combination

$$c_1 H_\lambda^1(z) + c_2 H_\lambda^2(z).$$

We observe, in addition, that the *Hankel functions* $H_\lambda^1(z)$ *and* $H_\lambda^2(z)$ *are determined uniquely, up to a factor not containing* z, *by their behavior at infinity and by equation* (2). For, if there were two linearly independent solutions of Bessel's equation having the described property, say for the upper half-plane, then every solution—and, in particular, $H_\lambda^2(z)$—would have to have this property. But this contradicts the fact just proved that $|H_\lambda^2(z)|$ increases without limit on the positive imaginary axis.

Finally, we consider the Hankel functions, for fixed $z \neq 0$, with respect to their *dependence on the parameter* λ. Since the integrand in (3) depends analytically on λ and the integrals converge uniformly in every finite λ-domain, it follows that *the Hankel functions are analytic functions of* λ; in fact, they are *integral transcendental functions.*

3. Bessel and Neumann Functions. The solutions of equation (2) which are real for real λ and z are of particular interest in physics. In order to find expressions for them, we write

$$\begin{aligned} H_\lambda^1(z) &= J_\lambda(z) + iN_\lambda(z), \\ H_\lambda^2(z) &= J_\lambda(z) - iN_\lambda(z), \end{aligned} \tag{4}$$

[1] Let ξ_0 be chosen in such a way that $\pi/2 + \xi_0$ is an integral multiple of $\pi/2a$. Then the integral in question becomes

$$\int_0^{n\pi/2a} \cosh b\xi \, e^{y\cos\xi} \cos a\xi \, d\xi$$

$$= \frac{1}{a} \int_0^{\pi/2} \cos\xi \left\{ \sum_{\nu=0}^{n-1} (-1)^\nu \cosh \frac{b}{a}\left(\xi + \nu\frac{\pi}{2}\right) e^{y\cos(1/a)[\xi+\nu\pi/2]} \right\} d\xi.$$

Here the first term of the sum dominates as y increases, since the exponent $\cos(\xi/a)$ is larger by at least the amount $1 - \cos(\pi/2a)$ than any of the following exponents $\cos\frac{1}{a}\left(\xi + \nu\frac{\pi}{2}\right)$; this term increases with y beyond all bounds.

where

$$J_\lambda(z) = \tfrac{1}{2}(H_\lambda^1(z) + H_\lambda^2(z)) \tag{5}$$

is called the *Bessel function* of index λ and

$$N_\lambda(z) = \frac{1}{2i}(H_\lambda^1(z) - H_\lambda^2(z)) \tag{5'}$$

the corresponding *Neumann function*. Since the determinant

$$\begin{vmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2i} & -\frac{1}{2i} \end{vmatrix} = \frac{i}{2}$$

of the substitution is different from zero, *the functions* $J_\lambda(z)$ *and* $N_\lambda(z)$ *are linearly independent for all* λ.

For real z *and real* λ *the Hankel functions* $H_\lambda^1(z)$ *and* $H_\lambda^2(z)$ *are complex conjugates of each other.* For, if we replace ζ by $-\zeta$ in the representation

$$\overline{H_\lambda^1(z)} = -\frac{1}{\pi}\int_{\bar{L}_1} e^{iz\sin\zeta - \bar{\lambda}\zeta}\, d\zeta,$$

where $\bar{L}_1$ is the reflection of L_1 across the real axis, we have

$$\overline{H_\lambda^1(z)} = \frac{1}{\pi}\int_{-\bar{L}_1} e^{-iz\sin\zeta + \bar{\lambda}\zeta}\, d\zeta;$$

since the path $-\bar{L}_1$ is the same as L_2 taken in the negative sense, we obtain

$$\overline{H_\lambda^1(z)} = -\frac{1}{\pi}\int_{L_2} e^{-iz\sin\zeta + \bar{\lambda}\zeta}\, d\zeta = H_\lambda^2(z).$$

Hence, $J_\lambda(z)$ is the real and $N_\lambda(z)$ the imaginary part of the Hankel function $H_\lambda^1(z)$ for real λ and z, and therefore $J_\lambda(z)$ and $N_\lambda(z)$ are real.

The function $H_{-\lambda}^\nu(z)$ $(\nu = 1, 2)$ is a solution of Bessel's equation for every value of λ for which $H_\lambda^\nu(z)$ is a solution, since only λ^2 appears in the differential equation. The functions $H_\lambda^\nu(z)$ and $H_{-\lambda}^\nu(z)$ cannot be linearly independent because, by subsection 2, they show the same behavior at infinity.

In fact, if we introduce the new variable of integration $-\zeta - \pi$ in the representation

$$H^1_{-\lambda}(z) = -\frac{1}{\pi}\int_{L_1} e^{-iz\sin\zeta - i\lambda\zeta}\, d\zeta,$$

we at once obtain the relation

$$H^1_{-\lambda}(z) = e^{i\lambda\pi} H^1_\lambda(z) \tag{6}$$

and, by means of a similar calculation,

$$H^2_{-\lambda}(z) = e^{-i\lambda\pi} H^2_\lambda(z). \tag{6'}$$

For the Bessel and Neumann functions of negative index one obtains

$$J_{-\lambda}(z) = \frac{e^{i\lambda\pi}H^1_\lambda(z) + e^{-i\lambda\pi}H^2_\lambda(z)}{2} \tag{7}$$

and

$$N_{-\lambda}(z) = \frac{e^{i\lambda\pi}H^1_\lambda(z) + e^{-i\lambda\pi}H^2_\lambda(z)}{2i}; \tag{7'}$$

unlike the Hankel functions, they are not linearly dependent on the functions $J_\lambda(z)$ and $N_\lambda(z)$, respectively, for every λ, but only for those values of λ for which the determinant

$$\frac{1}{4}\begin{vmatrix} e^{i\lambda\pi} & e^{-i\lambda\pi} \\ 1 & 1 \end{vmatrix} = \frac{i}{2}\sin\lambda\pi$$

vanishes, in other words, only when λ is an integer n. In this case we have

$$J_{-n}(z) = (-1)^n J_n(z) \tag{8}$$

and

$$N_{-n}(z) = (-1)^n N_n(z). \tag{8'}$$

The *general solution* of equation (2) can therefore be represented in the form

$$c_1 J_\lambda(z) + c_2 J_{-\lambda}(z)$$

when λ is not an integer n. On the other hand, if $\lambda = n$ we use the sum

$$c_1 J_n(z) + c_2 N_n(z).$$

It will be seen later, however, that even in this case $N_n(z)$ can be calculated easily from $J_n(z)$ and $J_{-n}(z)$ (see subsection 9).

4. Integral Representations of Bessel Functions. If we add the integrals (3) for $H_\lambda^1(z)$ and $H_\lambda^2(z)$, the paths of integration over the negative imaginary axis cancel and in the right half-plane $\mathcal{R}_e(z) > 0$ we obtain the representation

$$J_\lambda(z) = -\frac{1}{2\pi}\int_L e^{-iz\sin\zeta + i\lambda\zeta}\, d\zeta \tag{9}$$

for $J_\lambda(z)$, where L is the path indicated in Figure 11.

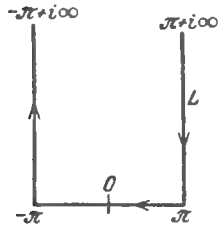


Figure 11

In particular, if $\lambda = n$ is an integer the integrals over the vertical portions of the path L also cancel, because of the periodicity of the integrand. It follows that

$$J_n(z) = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{iz\sin\zeta - in\zeta}\, d\zeta \tag{10}$$

or, since the real part of the integrand is an even function and the imaginary part is odd,

$$J_n(z) = \frac{1}{\pi}\int_0^{\pi} \cos(z\sin\zeta - n\zeta)\, d\zeta. \tag{10'}$$

By means of these integrals $J_n(z)$ is defined for all z. We see that *the Bessel functions with integral index are regular and single-valued in the whole plane and hence are integral functions.*

The representation (10) implies also that $J_n(z)$ is the n-th Fourier coefficient in the Fourier expansion of

$$e^{iz\sin\zeta} = \sum_{-\infty}^{\infty} J_n(z) e^{in\zeta} \tag{11}$$

with respect to ζ. This expansion can be considered as the definition of the function $J_n(z)$ for integral n by means of a *generating function* $e^{iz\sin\zeta}$.

For real z and ζ, (11) yields the relations

$$\cos(z \sin \zeta) = \sum_{-\infty}^{\infty} J_n(z) \cos n\zeta,$$

$$\sin(z \sin \zeta) = \sum_{-\infty}^{\infty} J_n(z) \sin n\zeta,$$

which also hold for complex z and ζ.

If we observe that

$$J_{-n}(z) = (-1)^n J_n(z)$$

we obtain

$$\begin{aligned} \cos(z \sin \zeta) &= J_0(z) + 2 \sum_{1}^{\infty} J_{2n}(z) \cos 2n\zeta, \\ \sin(z \sin \zeta) &= 2 \sum_{1}^{\infty} J_{2n-1}(z) \sin (2n-1)\zeta; \end{aligned} \tag{12}$$

in particular for $\zeta = \pi/2$ we have

$$\begin{aligned} \cos z &= J_0(z) - 2J_2(z) + 2J_4(z) - + \ldots, \\ \sin z &= 2J_1(z) - 2J_3(z) + - \ldots\ldots. \end{aligned}$$

If the new variable of integration $\zeta' = e^{-i\zeta}$ is introduced in (9), it follows that

$$J_\lambda(z) = \frac{1}{2\pi i} \int_L e^{z(\zeta - \zeta^{-1})/2} \zeta^{-\lambda-1} \, d\zeta, \tag{13}$$

where L is the path shown in Figure 12. It extends on both sides of the negative real axis up to the point $\zeta = -1$ and then circles the origin along the unit circle.[1]

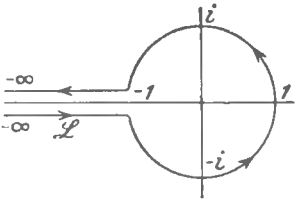


Figure 12

[1] We could have obtained this representation directly, on the basis of the method outlined in §1, by requiring the transformation kernel to satisfy the differential equation

$$z^2 K_{zz} + zK_z + z^2 K - \zeta(\zeta K_\zeta)_\zeta = 0,$$

which is solved by the function $K = e^{z(\zeta - \zeta^{-1})/2}$. The transformed differential equation is then $[\zeta(\zeta v)']' - \lambda^2 v = 0$; it has the solution $v = \zeta^{\pm\lambda-1}$.

For integral $\lambda = n$ the integrals on the straight portions cancel, and we have

$$J_n(z) = \frac{1}{2\pi i} \oint e^{z(\zeta - \zeta^{-1})/2} \zeta^{-n-1}\, d\zeta. \tag{14}$$

Thus $J_n(z)$ is the n-th coefficient in the Laurent expansion

$$e^{z(\zeta - \zeta^{-1})/2} = \sum_{-\infty}^{\infty} J_n(z)\zeta^n. \tag{15}$$

This expansion could also have been employed in the definition of $J_n(z)$ for integral n.

If we make the transformation $\zeta = 2v/z$ in (13)—at first under the assumption of real $z > 0$— taking the same path of integration we find

$$J_\lambda(z) = \frac{1}{2\pi i}\left(\frac{z}{2}\right)^\lambda \int_L e^{v - z^2/4v} v^{-(\lambda+1)}\, dv. \tag{16}$$

But since the integral on the right converges for all values of z, (16) furnishes a representation of the Bessel functions for all z. In particular, we see that *the quotient $J_\lambda(z)/z^\lambda$ is an integral function of z for every λ.*

5. Another Integral Representation of the Hankel and Bessel Functions. Another integral representation for the Bessel functions is obtained by considering the differential equation for $J_\lambda(z)/z^\lambda$ and applying the Laplace transformation. (Indeed, it seems reasonable to expect that in this way we shall arrive at particularly simple results, since $J_\lambda(z)/z^\lambda$ is a single-valued function of z.) For this purpose, we introduce in the equation

$$u'' + \frac{1}{z}u' + \left(1 - \frac{\lambda^2}{z^2}\right)u = 0$$

the new variable $\omega(z)$ by means of

$$u = \omega z^\lambda,$$

from which we obtain

$$z\omega'' + (2\lambda + 1)\omega' + z\omega = 0. \tag{17}$$

Writing

$$\omega(z) = \int_C K(z, \zeta)v(\zeta)\, d\zeta, \qquad K = e^{z\zeta},$$

we obtain

$$\int_C \{zK_{zz} + (2\lambda + 1)K_z + zK\}v(\zeta)\, d\zeta = 0.$$

In the special case of the Laplace transformation we have

$$K_z = \zeta K,$$

$$K_\zeta = zK,$$

and hence $zK_{zz} = \zeta^2 K_\zeta$; we therefore obtain

$$\int_C \{(1 + \zeta^2)K_\zeta + (2\lambda + 1)\zeta K\}v(\zeta)\, d\zeta$$

$$= -\int_C K(z, \zeta)\{(1 + \zeta^2)v' - (2\lambda - 1)\zeta v\}\, d\zeta$$

$$+ \int_C \frac{\partial}{\partial \zeta}(Kv(1 + \zeta^2))\, d\zeta = 0.$$

Thus the differential equation is solved if $v(\zeta)$ and C are determined in such a way that

$$(1 + \zeta^2)v'(\zeta) - (2\lambda - 1)\zeta v(\zeta) = 0$$

and $e^{z\zeta}v(\zeta)(1 + \zeta^2)$ takes equal values at the ends of C. From

$$\frac{v'(\zeta)}{v(\zeta)} = \frac{2\lambda - 1}{1 + \zeta^2}\zeta$$

we have

$$v(\zeta) = c(1 + \zeta^2)^{\lambda - 1/2}.$$

Thus

$$\omega(z) = c\int_C e^{z\zeta}(1 + \zeta^2)^{\lambda - 1/2}\, d\zeta$$

or, if we introduce $i\zeta$ as a new variable of integration, include $i(-1)^{\lambda - 1/2}$ among the constants, and denote the path of integration by C, we have

$$\omega(z) = c\int_C e^{iz\zeta}(\zeta^2 - 1)^{\lambda - 1/2}\, d\zeta.$$

In order to find an admissible path of integration, we first construct the Riemann surface of the integrand which is one-valued on the

Riemann surface of $\log(\zeta - 1) + \log(\zeta + 1) = \chi$. If we cut the ζ-plane along two rays parallel to the positive imaginary axis, starting at $\zeta = +1$ and $\zeta = -1$ and ending at $\zeta = +i\infty$, we obtain a simply connected domain in which a branch of χ is uniquely determined by the value of the imaginary part of χ on the positive real axis of the ζ-plane. We define two branches B_1 and B_2 of χ such that the imaginary part of χ is zero on B_1 and equals $2\pi i$ on B_2 for real positive values of ζ. A path crossing either of the branch cuts from the right-hand side to the left-hand side leads from B_1 into B_2.

Figure 13

Now let C_1 and C_2 denote paths lying in the branches B_1 and B_2, respectively, such that C_1 encloses the ray starting from $\zeta = +1$ and C_2 encloses the ray starting from $\zeta = -1$ without intersecting either of these rays (see Figure 13). Instead of introducing rays through $\zeta = \pm 1$ which are parallel to the imaginary axis, we can introduce rays making an angle α with the positive real axis (where $0 < \alpha < 2\pi$) and then define C_1 and C_2 as before.

The integral $\omega(z)$ converges on C_1 or C_2 for those z for which $\mathcal{R}_e(iz\zeta)$ tends to $-\infty$ along the ray. At the same time the expression

$$Kv(\zeta^2 - 1) = (\zeta^2 - 1)^{\lambda+1/2} e^{iz\zeta}$$

tends to zero at both ends of the path of integration; in other words, $\omega(z)$ is a solution of (17). If the ray makes the angle α with the ξ-axis, this will be the case when

$$y \cos\alpha + x \sin\alpha > 0$$

and thus when $z = x + iy$ lies in one of the half-planes bounded by the line $y \cos\alpha + x \sin\alpha = 0$. However, we can continue the integrals analytically, as in subsection 2, by letting α run through an unbounded sequence of positive values and a similar sequence of negative values in a suitable way. In particular, if for both paths we choose $\alpha = \pi/2$ as in Figure 13, both integrals converge in the right half-plane $\mathcal{R}_e(z) > 0$. If we rotate the path C_1 so that it lies along the positive real axis, then the corresponding integral converges in the upper half-plane and tends to zero as z increases without limit in the sector

$$\delta \leq \arg z \leq \pi - \delta \qquad (0 < \delta < \pi/2).$$

According to the observation of subsection 2, the integral must therefore coincide, up to a factor independent of z, with $H_\lambda^1(z)/z^\lambda$.

We find

$$H_\lambda^1(z) = a_1 z^\lambda \int_{C_1} e^{iz\zeta}(\zeta^2 - 1)^{\lambda-1/2}\, d\zeta.$$

and similarly

$$H_\lambda^2(z) = a_2 z^\lambda \int_{C_2} e^{iz\zeta}(\zeta^2 - 1)^{\lambda-1/2}\, d\zeta.$$

The coefficients a_1 and a_2, which can depend only on λ, are equal but have opposite signs. This follows—at first for real λ—from the remark that the Hankel functions are complex conjugates of each other for real λ and z (see subsection 3).

We find[1] for real values of λ and z:

$$\overline{H_\lambda^{(2)}(z)} = \bar{a}_2 z^\lambda \int_{C_1} e^{iz\zeta}(\zeta^2 - 1)^{\lambda-1/2}\, d\zeta.$$

This can be verified by observing that for large values of $|\zeta|$ in the proper branches of B_1 and B_2, the imaginary part of $\log(\zeta^2-1)$ on the four parallels to the imaginary axis in Figure 13 is approximately (i.e. for $|\zeta| \to \infty$ and for ζ approaching the branch cuts from different sides), from left to right, $-\pi i$, πi, $-\pi i$, πi. Therefore we have

$$\overline{H_\lambda^{(2)}(z)} = \frac{\bar{a}_2}{a_1} H_\lambda^{(1)}(z) = H_\lambda^{(1)},$$

that is $\bar{a}_2 = a_1$. Now we shall show that $a_2 = -a_1$ by proving that a_1 and a_2 are pure imaginary if λ is real and $\lambda > -\frac{1}{2}$. Since a_1 and a_2 depend analytically on λ, $a_1 + a_2 = 0$ for all values of λ. That a_1 is pure imaginary can be seen by taking $z = iy$, $y > 0$ and by turning the path C_1 into a path along the positive real axis from $\zeta = 1$ to $\zeta = +\infty$. Then

$$H_\lambda^{(1)}(iy) = a_1(1 - e^{-2\pi i(\lambda-1/2)})(iy)^\lambda \int_1^\infty e^{-y\zeta}(\zeta^2 - 1)^{\lambda-1/2}\, d\zeta.$$

From the formula

$$H_\lambda^1(iy) = \frac{e^{-\pi i\lambda/2}}{\pi i} \int_{-\infty}^{\infty} e^{-y\cosh\eta - \lambda\eta}\, d\eta$$

[1] In order to insure convergence on the positive real axis, we take C_1 and C_2 parallel to the imaginary axis (see Figure 13).

we find that

$$a_1(\lambda) \cos \pi\lambda = \frac{b(\lambda)}{\pi i}$$

where $b(\lambda)$ is real and positive for real $\lambda > -\frac{1}{2}$. For such λ we have, therefore, $\bar{a}_1 = -a_1$ so that $a_2 = -a_1$.

Now since by subsection 2 the Hankel functions and, as we see immediately, also the integrals $\int e^{iz\zeta}(\zeta^2 - 1)^{\lambda-1/2}\, d\zeta$ depend analytically on λ, the coefficients $a_1(\lambda)$ and $a_2(\lambda)$ are analytic functions of λ. In other words, the relation $a_1 = -a_2 = c$ holds generally.

Figure 14

If we add the two integral representations for the Hankel functions, we can deform the resulting path of integration into the figure eight $\mathfrak{A}$ of Figure 14, which circles $+1$ in the positive sense and -1 in the negative sense. We obtain a representation of the Bessel functions

$$J_\lambda(z) = cz^\lambda \int_{\mathfrak{A}} e^{iz\zeta}(\zeta^2 - 1)^{\lambda-1/2}\, d\zeta,$$

which, for $\lambda \neq n + \frac{1}{2}$ $(n = 0, \pm 1, \cdots)$, is valid in the entire z-plane, since the path of integration lies in the finite part of the plane.

To determine the constant c, we compare this representation with the integral representation (16) and find for $z = 0$ the relation

$$c \int_{\mathfrak{A}} (\zeta^2 - 1)^{\lambda-1/2}\, d\zeta = \frac{1}{2^\lambda} \frac{1}{2\pi i} \int_L e^v v^{-\lambda-1}\, dv.$$

As will be seen in the next subsection, the integral on the left has the value

$$\int_{\mathfrak{A}} (\zeta^2 - 1)^{\lambda-1/2}\, d\zeta = 2\pi i \frac{\Gamma(\frac{1}{2})}{\Gamma(\lambda + 1)\Gamma(\frac{1}{2} - \lambda)}.$$

In order to evaluate the integral on the right, we consider the integral

$$\frac{1}{2\pi i} \int_L e^v v^{t-1}\, dv,$$

in particular for positive real values of t. Since this integral represents an analytic function of t, it is sufficient to reduce it to known analytic functions for such t.

Under the assumption $t > 0$, we can contract the unit circle to the origin; since the exponent $t - 1$ is greater than -1, the integral is convergent up to the origin. By Cauchy's theorem, the value of the integral does not change if, instead of integrating over L, we integrate from $-\infty$ to 0 below the real axis and then integrate from 0 to $-\infty$ above the real axis (see Figure 15):

Figure 15

$$\frac{1}{2\pi i}\int_L e^v v^{t-1}\,dv = \underbrace{\frac{1}{2\pi i}\int_{-\infty}^{0} e^v v^{t-1}\,dv}_{\text{below the real axis}} + \underbrace{\frac{1}{2\pi i}\int_0^{-\infty} e^v v^{t-1}\,dv}_{\text{above the real axis}} \qquad (t > 0).$$

If we set $v = -w$, the first integral becomes

$$\frac{1}{2\pi i}\int_\infty^0 w^{t-1}e^{-(t-1)\pi i}e^{-w}(-dw) = \frac{1}{2\pi i}\int_0^\infty w^{t-1}e^{-(t-1)\pi i}e^{-w}\,dw,$$

the second becomes

$$\frac{1}{2\pi i}\int_0^\infty w^{t-1}e^{(t-1)\pi i}e^{-w}(-dw);$$

thus their sum is

$$\frac{1}{2\pi i}\int_0^\infty w^{t-1}e^{-w}(e^{t\pi i} - e^{-t\pi i})\,dw.$$

Since $e^{t\pi i} - e^{-t\pi i} = 2i \sin \pi t$ and, by definition,

$$\int_0^\infty w^{t-1}e^{-w}\,dw = \Gamma(t),$$

the sum of the integrals has the value

$$\frac{\sin \pi t}{\pi}\,\Gamma(t).$$

From the supplementary theorem for the gamma function,

$$\Gamma(t)\Gamma(1-t) = \frac{\pi}{\sin \pi t},$$

it follows that

$$\frac{\sin \pi t}{\pi}\,\Gamma(t) = \frac{1}{\Gamma(1-t)}.$$

We have, therefore,

$$\frac{1}{2\pi i}\int_L v^{t-1}e^v\,dv = \frac{1}{\Gamma(1-t)}.$$

Hence for the constant c we find the value

$$c = \frac{1}{2\pi i}\frac{1}{2^\lambda}\frac{\Gamma(\frac{1}{2}-\lambda)}{\Gamma(\frac{1}{2})}$$

and finally obtain for $J_\lambda(z)$ the representation

$$J_\lambda(z) = \frac{\Gamma(\frac{1}{2}-\lambda)}{2\pi i\Gamma(\frac{1}{2})}\left(\frac{z}{2}\right)^\lambda\int_{\mathfrak{A}} e^{iz\zeta}(\zeta^2-1)^{\lambda-1/2}\,d\zeta. \tag{18}$$

This representation holds for all λ except $\lambda = n + \frac{1}{2}$, where n is an integer ≥ 0.

Corresponding formulas are found for the Hankel functions:

$$\begin{aligned} H_\lambda^1(z) &= \frac{1}{\pi i}\frac{\Gamma(\frac{1}{2}-\lambda)}{\Gamma(\frac{1}{2})}\left(\frac{z}{2}\right)^\lambda\int_{C_1} e^{iz\zeta}(\zeta^2-1)^{\lambda-1/2}\,d\zeta, \\ H_\lambda^2(z) &= -\frac{1}{\pi i}\frac{\Gamma(\frac{1}{2}-\lambda)}{\Gamma(\frac{1}{2})}\left(\frac{z}{2}\right)^\lambda\int_{C_2} e^{iz\zeta}(\zeta^2-1)^{\lambda-1/2}\,d\zeta. \end{aligned} \tag{18'}$$

If $\mathcal{R}_e(\lambda) > -\frac{1}{2}$, one can derive from (18) the very useful representation

$$J_\lambda(z) = \frac{1}{\Gamma(\frac{1}{2})\Gamma(\lambda+\frac{1}{2})}\left(\frac{z}{2}\right)^\lambda\int_{-1}^{+1} e^{iz\zeta}(1-\zeta^2)^{\lambda-1/2}\,d\zeta. \tag{19}$$

Setting $\zeta = \sin\tau$, we find, for $\mathcal{R}_e(\lambda) > -\frac{1}{2}$,

$$J_\lambda(z) = \frac{1}{\Gamma(\frac{1}{2})\Gamma(\lambda+\frac{1}{2})}\left(\frac{z}{2}\right)^\lambda\int_{-\pi/2}^{+\pi/2}\cos(z\sin\tau)(\cos\tau)^{2\lambda}\,d\tau. \tag{20}$$

6. Power Series Expansion of Bessel Functions. One can obtain a power series expansion for $J(z)/z^\lambda$, which is single-valued and analytic in the entire z-plane, on the basis of elementary considerations. As in Chapter V, we could make the substitution

$$u(z) = z^\lambda\sum_0^\infty a_\nu z^\nu$$

in the differential equation (2) and successively determine the coefficients a_ν. However, in the framework of our present approach

the power series expansions will be obtained from integral representations.

We begin with the integral representation (18) and expand the function $e^{iz\zeta}$ in its power series; to be able to apply (18) we must assume that λ is not of the form $n + \frac{1}{2}$ ($n = 0, \pm 1, \pm 2, \cdots$). Since this series converges uniformly in every finite ζ-domain, we can integrate term by term and obtain

$$J_\lambda(z) = \frac{\Gamma(\frac{1}{2} - \lambda)}{2\pi i \Gamma(\frac{1}{2})} \left(\frac{z}{2}\right)^\lambda \sum_{n=0}^{\infty} \frac{z^n}{n!} i^n \int_{\mathfrak{A}} \zeta^n (\zeta^2 - 1)^{\lambda - 1/2}\, d\zeta.$$

In evaluating the integral $\int_{\mathfrak{A}} \zeta^n (\zeta^2 - 1)^{\lambda - 1/2}\, d\zeta$, we note that we deal with analytic functions of λ and that it is therefore sufficient to determine these functions for all λ such that $\mathcal{R}_e(\lambda) > 0$. In fact, in this case we can deform the path of integration into the interval $-1 \leq \lambda \leq 1$, traversed in both directions.

For the integrand we find the value

$$e^{\pi i(\lambda - 1/2)} \zeta^n (1 - \zeta^2)^{\lambda - 1/2} \text{ above the real axis,}$$

$$e^{-\pi i(\lambda - 1/2)} \zeta^n (1 - \zeta^2)^{\lambda - 1/2} \text{ below the real axis,}$$

and therefore

$$\int_{\mathfrak{A}} \zeta^n (\zeta^2 - 1)^{\lambda - 1/2}\, d\zeta = -2i \sin \pi(\lambda - \tfrac{1}{2}) \int_{-1}^{1} \zeta^n (1 - \zeta^2)^{\lambda - 1/2}\, d\zeta.$$

The integral on the right vanishes for odd n; for even values of n we find

$$\int_{\mathfrak{A}} \zeta^{2n} (\zeta^2 - 1)^{\lambda - 1/2}\, d\zeta = 4i \sin \pi(\lambda + \tfrac{1}{2}) \int_{0}^{1} \zeta^{2n} (1 - \zeta^2)^{\lambda - 1/2}\, d\zeta.$$

Using the transformation $\zeta^2 = u$ this becomes

$$\int_{\mathfrak{A}} \zeta^{2n} (\zeta^2 - 1)^{\lambda - 1/2}\, d\zeta = 2i \sin \pi(\lambda + \tfrac{1}{2}) \int_{0}^{1} u^{n - 1/2} (1 - u)^{\lambda - 1/2}\, du.$$

The integral on the right is an Euler integral of the first kind. From the well-known relation

$$B(p, q) = \int_0^1 x^{p-1} (1 - x)^{q-1}\, dx = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p + q)}$$

we obtain

$$\int_{\mathfrak{A}} \zeta^{2n}(\zeta^2 - 1)^{\lambda-1/2}\, d\zeta = 2i \sin \pi(\lambda + \tfrac{1}{2}) B(n + \tfrac{1}{2}, \lambda + \tfrac{1}{2})$$

$$= 2i \sin \pi(\lambda + \tfrac{1}{2}) \frac{\Gamma(n + \frac{1}{2})\Gamma(\lambda + \frac{1}{2})}{\Gamma(\lambda + n + 1)}.$$

But $\Gamma(x)\Gamma(1 - x) = \dfrac{\pi}{\sin \pi x}$, hence $\Gamma(\lambda + \frac{1}{2}) \sin \pi(\lambda + \frac{1}{2}) = \dfrac{\pi}{\Gamma(\frac{1}{2} - \lambda)}$.
Therefore we have

$$\int_{\mathfrak{A}} \zeta^{2n}(\zeta^2 - 1)^{\lambda-1/2}\, d\zeta = 2\pi i \frac{\Gamma(n + \frac{1}{2})}{\Gamma(\frac{1}{2} - \lambda)\Gamma(n + \lambda + 1)}$$

and in particular, for $n = 0$,

$$\int_{\mathfrak{A}} (\zeta^2 - 1)^{\lambda-1/2}\, d\zeta = 2\pi i \frac{\Gamma(\frac{1}{2})}{\Gamma(\lambda + 1)\Gamma(\frac{1}{2} - \lambda)}.$$

Substitution of the values obtained in our series for $J_\lambda(z)$ yields

$$J_\lambda(z) = \frac{1}{\Gamma(\frac{1}{2})} \left(\frac{z}{2}\right)^\lambda \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!} z^{2n} \frac{\Gamma(n + \frac{1}{2})}{\Gamma(n + \lambda + 1)}$$

or, since

$$\Gamma\left(n + \frac{1}{2}\right) = \frac{(2n)!}{2^{2n} n!} \Gamma\left(\frac{1}{2}\right),$$

we have

$$J_\lambda(z) = \left(\frac{z}{2}\right)^\lambda \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} \left(\frac{z}{2}\right)^{2n} \frac{1}{\Gamma(n + \lambda + 1)}. \tag{21}$$

If λ is not an integer, none of the coefficients $1/\Gamma(n + \lambda + 1)$ vanish. However, if λ is an integer, then

$$\frac{1}{\Gamma(n + \lambda + 1)} = 0 \qquad \text{for} \quad n + \lambda + 1 \leq 0,$$

$$\frac{1}{\Gamma(n + \lambda + 1)} = \frac{1}{(n + \lambda)!} \qquad \text{for} \quad n + \lambda + 1 > 0.$$

The above assumption $\lambda \neq n + \frac{1}{2}$ is seen to be unnecessary for the validity of the expansion (21), since the series (21) converges uniformly also for the values $\lambda = n + \frac{1}{2}$ and, as we have already seen, $J_\lambda(z)$ depends analytically on λ.

The series in (21) converges for all z and therefore $J_\lambda(z)/z^\lambda$ is an *integral transcendental function*, unless it is a polynomial or a constant. But the latter is impossible, since $\Gamma(n + \lambda + 1)$ is finite for negative integral λ in all but a finite number of cases. Hence the coefficient of z^{2n} cannot vanish, and the series for $J_\lambda(z)/z^\lambda$ always has infinitely many nonvanishing terms.

It is immediately apparent from the series expansion (21) that $J_\lambda(z)$ is real for real values of λ and z, since the gamma function has real values for real arguments.

7. Relations Between Bessel Functions. Now that we have derived power series expansions and integral representations for the Bessel functions, we shall develop a number of general properties of Bessel functions with the aid of the integral representation. We have equation (16)

$$J_\lambda(z) = \frac{1}{2\pi i}\left(\frac{z}{2}\right)^\lambda \int_L v^{-(\lambda+1)} e^{v - z^2/4v}\, dv,$$

in which L denotes the path of integration of Figure 12, page 475; therefore

$$\frac{J_\lambda(z)}{z^\lambda} = \frac{1}{2^\lambda}\frac{1}{2\pi i}\int_L v^{-(\lambda+1)} e^{v - z^2/4v}\, dv.$$

We differentiate with respect to z^2, taking the derivative on the right side under the integral sign in a formal way,

$$\frac{d^k}{d(z^2)^k}\frac{J_\lambda(z)}{z^\lambda} = \frac{1}{2^\lambda}\frac{1}{2\pi i}\int_L v^{-(\lambda+1)}\left(\frac{-1}{4v}\right)^k e^{v - z^2/4v}\, dv.$$

Differentiation under the integral sign is permissible since the inequality $|v| \geq 1$ holds on the path L and hence, for $|z| \leq h$, the function

$$|e^{-z^2/4v}| \leq e^{|z^2/4v|} \leq e^{h^2}$$

is uniformly bounded. The right side is therefore a uniformly convergent integral of an analytic function of z^2.

When the last equation is multiplied on both sides by $z^{\lambda+k}$ another Bessel function is obtained on the right, so that we have relation

$$\frac{d^k}{d(z^2)^k}\frac{J_\lambda(z)}{z^\lambda} = \left(-\frac{1}{2}\right)^k \frac{J_{\lambda+k}(z)}{z^{\lambda+k}} \tag{22}$$

or, in a different form,

$$\left(\frac{d}{z\,dz}\right)^k \frac{J_\lambda(z)}{z^\lambda} = (-1)^k \frac{J_{\lambda+k}(z)}{z^{\lambda+k}}.$$

In particular, for $k = 1$ we have

$$\frac{d}{dz}\frac{J_\lambda(z)}{z^\lambda} = -\frac{J_{\lambda+1}(z)}{z^\lambda}, \tag{23}$$

in other words, the recursion formula

$$\frac{dJ_\lambda(z)}{dz} = \frac{\lambda}{z} J_\lambda(z) - J_{\lambda+1}(z), \tag{24}$$

which for $\lambda = 0$ assumes the special form

$$J_1(z) = -\frac{dJ_0(z)}{dz}.$$

The cases $\lambda = -\frac{1}{2}$, $\lambda = \frac{1}{2}$ will be investigated further. By (21) we have

$$J_{-1/2}(z) = \left(\frac{z}{2}\right)^{-1/2} \sum_{n=0}^{\infty} \frac{(-1)^n}{n!\Gamma(n+\frac{1}{2})} \left(\frac{z}{2}\right)^{2n}$$

and, since

$$\begin{aligned}\Gamma(n+\tfrac{1}{2}) &= (n-\tfrac{1}{2})(n-\tfrac{3}{2})\cdots\tfrac{3}{2}\cdot\tfrac{1}{2}\Gamma(\tfrac{1}{2}) \\ &= (n-\tfrac{1}{2})(n-\tfrac{3}{2})\cdots\tfrac{3}{2}\cdot\tfrac{1}{2}\sqrt{\pi},\end{aligned}$$

this yields

$$J_{-\frac{1}{2}}(z) = \sqrt{\frac{2}{\pi z}} \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!} z^{2n} = \sqrt{\frac{2}{\pi z}} \cos z.$$

Using formula (23) for $\lambda = -\frac{1}{2}$,

$$\frac{d}{dz}\frac{J_{-1/2}(z)}{z^{-1/2}} = -\frac{J_{1/2}(z)}{z^{-1/2}},$$

we have

$$\frac{d}{dz}\sqrt{\frac{2}{\pi}}\cos z = -\sqrt{\frac{2}{\pi}}\sin z = -\frac{J_{1/2}(z)}{z^{-1/2}}$$

or, in other words,

$$J_{1/2}(z) = \sqrt{\frac{2}{\pi z}} \sin z; \tag{25}$$

dividing, we obtain

$$\frac{J_{-1/2}(z)}{J_{1/2}(z)} = \cot z.$$

Thus the Bessel functions for $\lambda = -\frac{1}{2}$, $\lambda = \frac{1}{2}$ may be expressed simply by trigonometric functions.

From the formulas

$$J_\lambda(z) = \frac{H_\lambda^1(z) + H_\lambda^2(z)}{2},$$

$$J_{-\lambda}(z) = \frac{H_\lambda^1(z)e^{i\lambda\pi} + H_\lambda^2(z)e^{-i\lambda\pi}}{2}$$

we obtain, for $\lambda = \frac{1}{2}$,

$$J_{1/2}(z) = \frac{H_{1/2}^1(z) + H_{1/2}^2(z)}{2}$$

and

$$J_{-1/2}(z) = \frac{i(H_{1/2}^1(z) - H_{1/2}^2(z))}{2}$$

or

$$-iJ_{-1/2}(z) = \frac{H_{1/2}^1(z) - H_{1/2}^2(z)}{2}.$$

Adding the two expressions we obtain

$$H_{1/2}^1(z) = J_{1/2}(z) - iJ_{-1/2}(z) = \sqrt{\frac{2}{\pi z}} (\sin z - i \cos z) \tag{26}$$

$$= -i \sqrt{\frac{2}{\pi z}} e^{iz},$$

subtracting,

$$H_{1/2}^2(z) = J_{1/2}(z) + iJ_{-1/2}(z) = \sqrt{\frac{2}{\pi z}} (\sin z + i \cos z) \tag{26'}$$

$$= i \sqrt{\frac{2}{\pi z}} e^{-iz}.$$

These formulas show again that *the relation between the Bessel, Neumann, and Hankel functions is similar to that between the sine, cosine, and exponential functions.* This analogy will be apparent also in connection with the theorems on the distribution of zeros which will be derived later (see subsection 8).

If we apply relation (22)

$$\frac{d^k}{d(z^2)^k}\frac{J_\lambda(z)}{z^\lambda} = \left(-\frac{1}{2}\right)^k \frac{J_{\lambda+k}(z)}{z^{\lambda+k}}$$

derived on page 485 to the case $\lambda = \frac{1}{2}$,

$$\frac{d^k}{d(z^2)^k}\sqrt{\frac{2}{\pi}}\frac{\sin z}{z} = \left(-\frac{1}{2}\right)^k \frac{J_{k+1/2}(z)}{z^{k+1/2}},$$

it follows that

$$J_{k+1/2}(z) = (-1)^k \frac{(2z)^{k+1/2}}{\sqrt{\pi}} \frac{d^k}{d(z^2)^k}\frac{\sin z}{z};$$

in other words, every Bessel function $J_{k+1/2}(z)$ may be expressed as a rational function of trigonometric functions and of z, multiplied by $\sqrt{z}$.

A different recursion formula is obtained by differentiating under the integral sign in the relation

$$J_\lambda(z) = \frac{1}{2\pi i}\int_L e^{z(\zeta-\zeta^{-1})/2}\zeta^{-\lambda-1}\,d\zeta.$$

This yields

$$J'_\lambda(z) = \frac{1}{4\pi i}\left\{\int_L e^{z(\zeta-\zeta^{-1})/2}\zeta^{-\lambda}\,d\zeta - \int_L e^{z(\zeta-\zeta^{-1})/2}\zeta^{-\lambda-2}\,d\zeta\right\},$$

$$J'_\lambda(z) = \tfrac{1}{2}\{J_{\lambda-1}(z) - J_{\lambda+1}(z)\}. \tag{27}$$

If we subtract from this the expression

$$J'_\lambda(z) = \frac{\lambda}{z}J_\lambda(z) - J_{\lambda+1}(z),$$

it follows that

$$J_{\lambda-1}(z) + J_{\lambda+1}(z) = \frac{2\lambda}{z}J_\lambda(z). \tag{28}$$

The last relation can also be written as

$$\frac{J_{\lambda-1}(z)}{J_\lambda(z)} = \frac{2\lambda}{z} - \frac{1}{\dfrac{J_\lambda(z)}{J_{\lambda+1}(z)}} = \frac{2\lambda}{z} - \cfrac{1}{\dfrac{2\lambda+2}{z} - \cfrac{1}{\dfrac{2\lambda+4}{z} - \ddots}}.$$

$J_{\lambda-1}(z)/J_\lambda(z)$ is thus represented by an infinite continued fraction; however, we cannot investigate the question of its convergence here. If we multiply through by z, the continued fraction assumes the form

$$z\frac{J_{\lambda-1}(z)}{J_\lambda(z)} = 2\lambda - \cfrac{z^2}{2\lambda+2 - \cfrac{z^2}{2\lambda+4 - \ddots}}. \tag{29}$$

For $\lambda = \frac{1}{2}$ this reduces to

$$z\frac{J_{-1/2}(z)}{J_{1/2}(z)} = z\cot z = 1 - \cfrac{z^2}{3 - \cfrac{z^2}{5 - \ddots}}, \tag{30}$$

which represents an *infinite continued fraction for* cot z, known in the 18th century and used by Lambert[1] to prove the irrationality of π, for which purpose he set $z = \pi/4$.

The following *functional relation* holds for Bessel functions of integral index n:

$$J_n(a+b) = \sum_{\nu=-\infty}^{\infty} J_\nu(a)J_{n-\nu}(b). \tag{31}$$

The proof is obtained directly by considering the generating function $e^{i(a+b)\sin\zeta} = e^{ia\sin\zeta}\cdot e^{ib\sin\zeta}$. We have accordingly

$$\sum_{n=-\infty}^{\infty} J_n(a+b)e^{in\zeta} = \sum_{n=-\infty}^{\infty}\left(\sum_{\nu=-\infty}^{\infty} J_\nu(a)J_{n-\nu}(b)\right)e^{in\zeta},$$

from which the assertion follows.

There is a generalization of this formula which, for $n = 0$, reads:

$$J_0(\sqrt{a^2+b^2+2ab\cos\alpha}) = J_0(a)J_0(b) + 2\sum_{1}^{\infty} J_\nu(a)J_{-\nu}(b)\cos\nu\alpha. \tag{32}$$

[1] J. H. Lambert, Mémoire sur quelques propriétés remarquables des quantités transcendentes circulaires et logarithmiques, Acad. sc. Berlin, Mém., Vol. 17 (1761), 1768, pp. 265–322, in particular p. 269.

To prove it, we make use of the integral representation (10) and write the product $J_\nu(a)J_{-\nu}(b)$ as a double integral,

$$\frac{1}{4\pi^2}\int_{-\pi}^{\pi} d\zeta_1 \int_{-\pi}^{\pi} e^{i(a\sin\zeta_1+b\sin\zeta_2)-i\nu(\zeta_1-\zeta_2)}\, d\zeta_2 .$$

With a little manipulation this can be brought into the form

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} J_0(\sqrt{a^2+b^2+2ab\cos\alpha})\, e^{-i\nu\alpha}\, d\alpha,$$

which proves relation (32).

Finally, we note that a function $f(r)$, subject to certain assumptions, can be represented with the aid of Bessel functions in a way similar to that in which it can be represented (according to Fourier's integral theorem) in terms of exponential functions (see Ch. II, §6 and Ch. V, §12). *Let $f(r)$ be continuous and piecewise smooth, and let*

$$\int_0^\infty r\,|\,f(r)\,|\,dr < \infty .$$

Then for every integer n and $r > 0$, $f(r)$ is represented by formula

$$f(r) = \int_0^\infty s\,ds \int_0^\infty tf(t)J_n(st)J_n(sr)\,dt. \tag{33}$$

This formula is derived in the following way: We set

$$x = r\cos\theta, \qquad y = r\sin\theta$$

and consider the function

$$g(x, y) = f(r)e^{in\vartheta},$$

which, under the above assumptions on $f(r)$, is certainly continuous and—except in the neighborhood of the origin—has piecewise continuous derivatives. If we apply Fourier's integral theorem for two dimensions (see Ch. II, §6, 2) to $g(x, y)$ and interchange the order of integration of the two inner integrals—which requires justification—we obtain

$$g(x, y) = \frac{1}{4\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{i(ux+vy)}\,du\,dv \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(\xi, \eta)e^{-i(u\xi+v\eta)}\,d\xi\,d\eta.$$

We now introduce polar coordinates for the variables of integration $\xi, \eta; u, v$:

$$\xi = s\cos\alpha, \qquad u = t\cos\beta,$$
$$\eta = s\sin\alpha, \qquad v = t\sin\beta,$$

and obtain

$$f(r)e^{in\theta} = \frac{1}{4\pi^2}\int_0^\infty t\,dt\int_{-\pi}^{\pi} e^{irt\cos(\beta-\theta)}\,d\beta\int_0^\infty sf(s)\,ds\int_{-\pi}^{\pi} e^{in\alpha}e^{-ist\cos(\alpha-\beta)}\,d\alpha.$$

If we make the substitution

$$\beta - \theta = \frac{\pi}{2} + \beta',$$

$$\alpha - \beta = \alpha' - \frac{\pi}{2}$$

this expression becomes

$$f(r)e^{in\theta} = \frac{e^{in\theta}}{4\pi^2}\int_0^\infty t\,dt\int_{-\pi}^{\pi} e^{-irt\sin\beta'+in\beta'}\,d\beta'\int_0^\infty sf(s)\,ds\int_{-\pi}^{\pi} e^{-ist\sin\alpha'+in\alpha'}\,d\alpha',$$

since the exponential functions are periodic. The asserted relation

$$f(r) = \int_0^\infty tJ_n(rt)\,dt\int_0^\infty sf(s)J_n(st)\,ds$$

follows immediately if we apply formula (10) and integrate with respect to α' and β'.

Instead of justifying the inversion of the order of integration, a proof similar to that for the Fourier integral theorem can be given for the integral formula (33). We employ relation

$$f(r) = \lim_{v\to\infty}\int_0^a sf(s)P_v(s,r)\,ds,$$
$$P_v(s,r) = \int_0^v tJ_n(st)J_n(rt)\,dt, \tag{34}$$

which holds for every piecewise smooth function $f(r)$ vanishing at the origin; a denotes any positive number $a > r > 0$. This relation corresponds exactly to the Dirichlet integral of Chapter II and is proved in a similar way. We shall show that, under the assumption that the integral $\int_0^\infty r\,|\,f(r)\,|\,dr$ exists, the integration with respect

to s can be extended to infinity. For, the identity

$$(35) \qquad P_v(r, s) = \frac{v}{s^2 - r^2}\{sJ_n(vr)J_{n+1}(vs) - rJ_n(vs)J_{n+1}(vr)\},$$

valid for $r \neq s$, which is obtained from (36) of the next subsection using recursion formula (24), implies that, for fixed $r \neq 0$, $P_v(r, s)$ tends to zero with increasing s uniformly in v (e.g. using the asymptotic expansions of the Bessel functions for large arguments—see Ch. V, §11, 2 or Ch. VII, §6, 2). Hence the integral

$$\int_a^b sf(s)P_v(r, s)\, ds$$

also becomes arbitrarily small, uniformly in v and b, for sufficiently large a. Our assertion—relation (33)—follows.

8. Zeros of Bessel Functions. Finally, we shall derive some theorems concerning the zeros of the Bessel functions[1].

The Bessel function $J_\lambda(z)$ satisfies the differential equation

$$J_\lambda''(z) + \frac{1}{z}J_\lambda'(z) + \left(1 - \frac{\lambda^2}{z^2}\right)J_\lambda(z) = 0.$$

If we set

$$z = \xi_1 t, \qquad \xi_1 = \text{const.} \neq 0,$$

we obtain

$$J_\lambda''(\xi_1 t) + \frac{1}{\xi_1 t}J_\lambda'(\xi_1 t) + \left(1 - \frac{\lambda^2}{\xi_1^2 t^2}\right)J_\lambda(\xi_1 t) = 0.$$

Similarly, for

$$z = \xi_2 t, \qquad \xi_2 = \text{const.} \neq 0,$$

we have

$$J_\lambda''(\xi_2 t) + \frac{1}{\xi_2 t}J_\lambda'(\xi_2 t) + \left(1 - \frac{\lambda^2}{\xi_2^2 t^2}\right)J_\lambda(\xi_2 t) = 0.$$

If we multiply the first of these equations by $\xi_1^2 t J_\lambda(\xi_2 t)$, the second by $-\xi_2^2 t J_\lambda(\xi_1 t)$, and add the resulting expressions, we obtain

[1] Cf. the related results in Ch. VI, §2, 4.

$$t(\xi_1^2 J_\lambda''(\xi_1 t)J_\lambda(\xi_2 t) - \xi_2^2 J_\lambda''(\xi_2 t)J_\lambda(\xi_1 t))$$
$$+ (\xi_1 J_\lambda'(\xi_1 t)J_\lambda(\xi_2 t) - \xi_2 J_\lambda'(\xi_2 t)J_\lambda(\xi_1 t))$$
$$+ (\xi_1^2 - \xi_2^2)tJ_\lambda(\xi_1 t)J_\lambda(\xi_2 t) = 0.$$

The first two terms taken together are equal to the derivative of the function

$$t(\xi_1 J_\lambda'(\xi_1 t)J_\lambda(\xi_2 t) - \xi_2 J_\lambda'(\xi_2 t)J_\lambda(\xi_1 t))$$

with respect to t. If this function exists over the interval (0, 1), we obtain, integrating the last equation from 0 to 1,

$$t(\xi_1 J_\lambda'(\xi_1 t)J_\lambda(\xi_2 t) - \xi_2 J_\lambda'(\xi_2 t)J_\lambda(\xi_1 t))\,|_0^1$$
$$+ (\xi_1^2 - \xi_2^2)\int_0^1 tJ_\lambda(\xi_1 t)J_\lambda(\xi_2 t)\,dt = 0. \tag{36}$$

Since $\dfrac{J_\lambda(z)}{z^\lambda}$ is an entire function of z it is not difficult to show that

$$t(\xi_1 J_\lambda'(\xi_1 t)J_\lambda(\xi_2 t) - \xi_2 J_\lambda'(\xi_2 t)J_\lambda(\xi_1 t)) = t^{2\lambda+2}g(t),$$

where $g(t)$ is an entire function of t. Therefore, if $\lambda > -1$, this expression vanishes at $t = 0$ and (36) leads to

$$(\xi_1 J_\lambda'(\xi_1)J_\lambda(\xi_2) - \xi_2 J_\lambda'(\xi_2)J_\lambda(\xi_1))$$
$$+ (\xi_1^2 - \xi_2^2)\int_0^1 tJ_\lambda(\xi_1 t)\,J_\lambda(\xi_2 t)dt = 0. \tag{37}$$

From this equation, conclusions may be drawn about the distribution of the zeros of $J_\lambda(z)$ (see Ch. VI, §6).

Let ξ be a zero of $J_\lambda(z)$ which is different from zero. We set $\xi_1 = \xi$, $\xi_2 = \bar{\xi}$, where $\bar{\xi}$ denotes the complex conjugate of ξ. Thus ξ_1 and ξ_2 coincide only for real ξ.

Let λ be real, so that $J_\lambda(z)$ takes on real values for real z. The coefficients of the power series (21) are real; hence if $J_\lambda(\xi)$ vanishes, then $J_\lambda(\bar{\xi})$ also vanishes. If in equation (37) we set $J_\lambda(\xi_1) = J_\lambda(\xi_2) = 0$ the expression in large parentheses vanishes, and the second term becomes

$$(\xi^2 - \bar{\xi}^2)\int_0^1 t\,|\,J_\lambda(\xi t)\,|^2\,dt = 0.$$

It was assumed that $\xi \neq 0$. Since the Bessel function does not vanish identically, $\int_0^1 t \mid J_\lambda(\xi t) \mid^2 dt \neq 0$ and hence $\xi^2 - \bar{\xi}^2 = (\xi - \bar{\xi})(\xi + \bar{\xi}) = 0$, which implies

$$\xi = \bar{\xi} \quad \text{or} \quad \xi = -\bar{\xi}.$$

Hence ξ is either real or pure imaginary. For real $\lambda > -1$ the Bessel function $J_\lambda(z)$ has only real or pure imaginary zeros.

In order to investigate the pure imaginary zeros of the Bessel functions, we consider the power series expansion

$$\frac{J_\lambda(z)}{z^\lambda} = \frac{1}{2^\lambda} \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} \left(\frac{z}{2}\right)^{2n} \frac{1}{\Gamma(n + \lambda + 1)}.$$

The substitution $z = ai$, a real $\neq 0$, yields

$$\frac{J_\lambda(z)}{z^\lambda} = \frac{1}{2^\lambda} \sum_{n=0}^{\infty} \left(\frac{a}{2}\right)^{2n} \frac{1}{n!\Gamma(n + \lambda + 1)}.$$

Since λ is real, $n + \lambda + 1$ is positive for all n with a finite number of exceptions, and since the gamma function takes on positive values for positive arguments, all the coefficients of the power series are positive except perhaps for a finite number at the beginning of the series. For large $\mid a \mid$ the higher powers prevail, and since $(a/2)^{2n} > 0$ for $a \neq 0$, we have $J_\lambda(z)/z^\lambda > 0$ for all sufficiently large $\mid a \mid$. Thus zeros of the function $J_\lambda(z)/z^\lambda$ can occur only on a finite piece of the imaginary axis; hence, as an integral transcendental function, $J_\lambda(z)/z^\lambda$ can have only a finite number of pure imaginary zeros. $J_\lambda(z)/z^\lambda$ has no pure imaginary zeros for $\lambda > -1$, since in this case for all n

$$n + \lambda + 1 > 0,$$

$$\Gamma(n + \lambda + 1) > 0.$$

Thus all coefficients of the series are positive, and the value of the series itself is positive. In particular for $\lambda = 0, 1, 2, \cdots$ there are no pure imaginary zeros.

We have the result: *For real* $\lambda > -1$, $J_\lambda(z)$ *has only real zeros. For any real value of* λ, $J_\lambda(z)$ *has only a finite number of pure imaginary zeros. Actually it can be proved that, for real* λ, *the total number of zeros not on the real axis is always finite.*

The function $J_\lambda(z)$ has infinitely many real zeros; for every real

positive integer λ this is clear from the discussion of the preceding chapter, since the zeros of $J_n(z)$ furnish the system of the eigenvalues of a differential equation.

We add a few remarks about the position of the real zeros of Bessel functions.

Suppose λ is real and set

$$\frac{J_\lambda(z)}{z^\lambda} = v, \qquad J_\lambda(z) = vz^\lambda;$$

then we have (17)

$$zv'' + (2\lambda + 1)v' + zv = 0.$$

If ξ is a positive zero of v', for $z = \xi$ the differential equation takes the form

$$\xi v''(\xi) + \xi v(\xi) = 0,$$

and hence

$$v''(\xi) + v(\xi) = 0.$$

It follows that at the point ξ the second derivative $v''(\xi)$ cannot also vanish. For, if it did, we would have $v(\xi) = 0$, from which, together with $v'(\xi) = 0$, the identical vanishing of the solution $v(z)$ of (17) would follow. We therefore conclude that $v(\xi)$ and $v''(\xi)$ are of opposite sign.

Now let ξ_1 and $\xi_2 (> \xi_1)$ be two adjacent positive zeros of $v'(z)$, so that $v'(z) \neq 0$ for $\xi_1 < z < \xi_2$. By Rolle's theorem there must then be an odd number of zeros of v'' between ξ_1 and ξ_2; consequently $v''(\xi_1)$ and $v''(\xi_2)$, and hence also $v(\xi_1)$ and $v(\xi_2)$, are of opposite sign. Therefore an odd number of zeros of v must lie between ξ_1 and ξ_2, hence at least one; but by Rolle's theorem there can be only one since an odd number of zeros of v' lies between two adjacent zeros of v, and by hypothesis v' has no zeros between ξ_1 and ξ_2. Accordingly, v has precisely one zero between ξ_1 and ξ_2. In other words, between two adjacent positive zeros of v' there is one and only one zero of v. *The positive zeros of v and v' separate each other, the same is true of the negative zeros.*

In subsection 7 we derived relation (23)

$$\frac{d}{dz}\frac{J_\lambda(z)}{z^\lambda} = -\frac{J_{\lambda+1}(z)}{z^\lambda}$$

or

$$v' = -\frac{J_{\lambda+1}(z)}{z^{\lambda}}.$$

Since the zeros of v and v' separate each other, and since moreover

$$v = \frac{J_{\lambda}(z)}{z^{\lambda}}, \qquad v' = -\frac{J_{\lambda+1}(z)}{z^{\lambda}},$$

and hence all positive and negative zeros of v and v' are also zeros of $J_{\lambda}(z)$ and $J_{\lambda+1}(z)$, we find in addition: *The zeros of* $J_{\lambda}(z)$ *and* $J_{\lambda+1}(z)$ *separate each other.*

For $\lambda = -\frac{1}{2}, \lambda = \frac{1}{2}$ we found that

$$J_{-\frac{1}{2}}(z) = \sqrt{\frac{2}{\pi z}}\cos z, \qquad J_{\frac{1}{2}}(z) = \sqrt{\frac{2}{\pi z}}\sin z.$$

The zeros of these functions are given by

$$\pm\frac{\pi}{2}, \qquad \pm\frac{3\pi}{2}, \qquad \pm\frac{5\pi}{2}, \qquad \cdots, \qquad \pm\frac{(2n+1)\pi}{2}, \cdots$$

and

$$0, \qquad \pm\pi, \qquad \pm 2\pi, \qquad \cdots, \qquad \pm n\pi, \qquad \cdots,$$

respectively; thus they do, in fact, separate each other.

In this respect, too, the Bessel functions reveal a relationship to trigonometric functions.

9. Neumann Functions. If λ is not an integer, relations (5)

$$J_{\lambda}(z) = \tfrac{1}{2}\left(H^1_{\lambda}(z) + H^2_{\lambda}(z)\right)$$

and (7)

$$J_{-\lambda}(z) = \tfrac{1}{2}\left(e^{i\lambda\pi}H^1_{\lambda}(z) + e^{-i\lambda\pi}H^2_{\lambda}(z)\right)$$

may be solved for $H^1_{\lambda}(z)$ and $H^2_{\lambda}(z)$. We obtain

$$H^1_{\lambda}(z) = -\frac{1}{i\sin\lambda\pi}\left(J_{\lambda}(z)e^{-i\lambda\pi} - J_{-\lambda}(z)\right), \tag{38}$$

$$H^2_{\lambda}(z) = \frac{1}{i\sin\lambda\pi}\left(J_{\lambda}(z)e^{i\lambda\pi} - J_{-\lambda}(z)\right), \tag{38'}$$

and hence

$$N_{\lambda}(z) = \frac{1}{2i}\left(H^1_{\lambda}(z) - H^2_{\lambda}(z)\right) = \frac{J_{\lambda}(z)\cos\lambda\pi - J_{-\lambda}(z)}{\sin\lambda\pi}. \tag{39}$$

This representation of the Neumann functions by means of J_λ and $J_{-\lambda}$ fails, however, if λ is an integer. For, considered as functions of λ, both the numerator $J_\lambda \cos \lambda\pi - J_{-\lambda}$ and the denominator $\sin \lambda\pi$ have simple zeros in this case. But since for $z \neq 0$ both numerator and denominator are regular analytic functions of λ, we may differentiate them both in order to determine the value of the function for integral λ. Passing to the limit, we find from the quotient

$$\frac{\dfrac{\partial J_\lambda(z)}{\partial \lambda} \cos \lambda\pi - J_\lambda(z)\pi \sin \lambda\pi - \dfrac{\partial J_{-\lambda}(z)}{\partial \lambda}}{\pi \cos \lambda\pi}$$

that for integral λ

$$N_\lambda(z) = \frac{1}{\pi}\left(\frac{\partial J_\lambda(z)}{\partial J} - (-1)^\lambda \frac{\partial J_{-\lambda}(z)}{\partial J}\right) \tag{40}$$

holds. It is easy to verify directly that the expression just obtained represents a solution of the differential equation for integral λ. In fact, if we differentiate Bessel's eqution

$$\frac{d^2 J_\lambda(z)}{dz^2} + \frac{1}{z}\frac{dJ_\lambda(z)}{dz} + \left(1 - \frac{\lambda^2}{z^2}\right) J_\lambda(z) = 0,$$

which represents an identity in λ, with respect to λ, we obtain

$$\frac{d^2}{dz^2}\frac{\partial J_\lambda(z)}{\partial \lambda} + \frac{1}{z}\frac{d}{dz}\frac{\partial J_\lambda(z)}{\partial \lambda} + \left(1 - \frac{\lambda^2}{z^2}\right)\frac{\partial J_\lambda(z)}{\partial \lambda} = \frac{2\lambda}{z^2} J_\lambda(z).$$

Similarly, for $-\lambda$ we find

$$\frac{d^2}{dz^2}\frac{\partial J_{-\lambda}(z)}{\partial \lambda} + \frac{1}{z}\frac{d}{dz}\frac{\partial J_{-\lambda}(z)}{\partial \lambda} + \left(1 - \frac{\lambda^2}{z^2}\right)\frac{\partial J_{-\lambda}(z)}{\partial \lambda} = \frac{2\lambda}{z^2} J_{-\lambda}(z).$$

We multiply the second equation by $(-1)^\lambda$ and subtract the resulting expression from the first equation; from the relation $J_\lambda(z) = (-1)^\lambda J_{-\lambda}(z)$ we see that the right side of the resulting equation vanishes for integral λ, and we obtain, as a further solution of Bessel's equation, the function

$$\frac{\partial J_\lambda(z)}{\partial \lambda} - (-1)^\lambda \frac{\partial J_{-\lambda}(z)}{\partial \lambda} = \pi N_\lambda(z) \qquad \text{(integral } \lambda\text{)}.$$

The relations just derived between the Neumann function $N_\lambda(z)$

and the functions $J_\lambda(z)$ and $J_{-\lambda}(z)$ enable us to find corresponding representations for $N_\lambda(z)$ from the representations for the Bessel functions. Thus, for example, one finds from the integral representation (9) for $\lambda \neq n$,

$$(41) \quad N_\lambda(z) = -\frac{1}{2\pi \sin \pi\lambda} \int_L e^{-iz\sin\zeta} \{e^{i\lambda\zeta} \cos \pi\lambda - e^{-i\lambda\zeta}\}\, d\zeta,$$

and for $\lambda = n$,

$$(42) \quad N_n(z) = -\frac{i}{\pi^2} \int_L \zeta e^{-iz\sin\zeta} \cos n\zeta\, d\zeta \qquad (n \text{ even}),$$

$$(42') \quad N_n(z) = \frac{1}{\pi^2} \int_L \zeta e^{-iz\sin\zeta} \sin n\zeta\, d\zeta \qquad (n \text{ odd}).$$

Making use of the integral formula (20) of subsection 5, it follows, e.g. for $N_0(z)$, in virtue of

$$N_0(z) = \frac{2}{\pi} \left(\frac{\partial J_\lambda}{\partial \lambda}\right)_{\lambda=0}$$

that

$$(43) \quad \pi N_0(z) = 2(C + \log 2) J_0(z) + \frac{2}{\pi} \int_0^\pi \cos(z \cos\zeta) \log(z \sin^2 \zeta)\, d\zeta,$$

where C is the well-known Euler constant.

Similarly, series expansions for $N_\lambda(z)$ may be obtained from those for $J_\lambda(z)$ and $J_{-\lambda}(z)$. We shall consider in greater detail the case of integral λ. We have (21)

$$J_\lambda(z) = \left(\frac{z}{2}\right)^\lambda \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} \left(\frac{z}{2}\right)^{2n} \frac{1}{\Gamma(n + \lambda + 1)}$$

and, since we may differentiate with respect to λ under the summation sign,

$$\frac{\partial J_\lambda(z)}{\partial \lambda} = \log \frac{z}{2} J_\lambda(z) + \left(\frac{z}{2}\right)^\lambda \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} \left(\frac{z}{2}\right)^{2n} \left(\frac{d}{dt} \frac{1}{\Gamma(t)}\right)_{t=n+\lambda+1},$$

$$(44) \quad \frac{\partial J_{-\lambda}(z)}{\partial \lambda} = -\log \frac{z}{2} J_{-\lambda}(z) - \left(\frac{z}{2}\right)^{-\lambda} \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} \left(\frac{z}{2}\right)^{2n} \left(\frac{d}{dt} \frac{1}{\Gamma(t)}\right)_{t=n-\lambda+1}$$

First we determine the values of the derivatives $d[1/\Gamma(t)]/dt$ for positive integers t. The gamma function satisfies the functional equation

$$\Gamma(t+1) = t\Gamma(t) \qquad \text{for } t \neq 0, -1, -2, \cdots.$$

By logarithmic differentiation we find

$$\frac{\Gamma'(t+1)}{\Gamma(t+1)} = \frac{1}{t} + \frac{\Gamma'(t)}{\Gamma(t)}$$

and, repeating k times,

$$\frac{\Gamma'(t+k+1)}{\Gamma(t+k+1)} = \frac{1}{t+k} + \frac{1}{t+k-1} + \cdots + \frac{1}{t} + \frac{\Gamma'(t)}{\Gamma(t)}$$

$$(k = 0, 1, 2, \cdots).$$

We have

$$\frac{d}{dt}\frac{1}{\Gamma(t)} = -\frac{\Gamma'(t)}{\Gamma^2(t)} = -\frac{1}{\Gamma(t)}\frac{\Gamma'(t)}{\Gamma(t)};$$

if we set $t = 1$, $k = n - 1$ we obtain the above formula

$$\frac{\Gamma'(n+1)}{\Gamma(n+1)} = \frac{1}{n} + \frac{1}{n-1} + \cdots + 1 + \frac{\Gamma'(1)}{\Gamma(1)}$$

$$= \frac{1}{n} + \frac{1}{n-1} + \cdots + 1 - C$$

for $n = 1, 2, 3, \cdots$. Knowing the value of $\Gamma'(t)/\Gamma(t)$ for the positive integers, we also know the desired value of the derivative $d[1/\Gamma(t)]/dt$ at these points:

$$\left(\frac{d}{dt}\frac{1}{\Gamma(t)}\right)_{t=1} = C;$$

$$\frac{d}{dt}\frac{1}{\Gamma(t)} = -\frac{1}{(t-1)!}\left\{\frac{1}{t-1} + \frac{1}{t-2} + \cdots + 1 - C\right\}$$

$$\text{for } t = 2, 3, \cdots.$$

In order to determine in addition the value of the derivative for negative integral values t, we solve the equation

$$\frac{\Gamma'(t+k+1)}{\Gamma(t+k+1)} = \frac{1}{t+k} + \frac{1}{t+k-1} + \cdots + \frac{1}{t} + \frac{\Gamma'(t)}{\Gamma(t)}$$

for $\Gamma'(t)/\Gamma(t)$. Multiplying by $-1/\Gamma(t)$ we obtain

$$-\frac{1}{\Gamma(t)}\frac{\Gamma'(t)}{\Gamma(t)} = \frac{1}{\Gamma(t)}\left\{\frac{1}{t} + \frac{1}{t+1} + \cdots + \frac{1}{t+k-1} + \frac{1}{t+k}\right\} - \frac{1}{\Gamma(t)}\frac{\Gamma'(t+k+1)}{\Gamma(t+k+1)}.$$

If we let t converge to $-k$ ($k = 0, 1, 2, \cdots$), the left side, and hence also the right side, converges to the derivative $(d[1/\Gamma(t)]/dt)_{t=-k}$. Now as $t \to -k$ the expression $1/\Gamma(t)$ tends to zero, and since

$$1/t + 1/(t+1) + \cdots + 1/(t+k-1)$$

and

$$\Gamma'(t+k+1)/\Gamma(t+k+1)$$

remain finite, all that is left on the right is the term $1/\Gamma(t)(t+k)$. If we multiply top and bottom by $t(t+1)\cdots(t+k-1)$ and use the functional equation of the gamma function, we see that the denominator becomes equal to $\Gamma(t+k+1)$ and converges to $\Gamma(1) = 1$ as $t \to -k$; since the numerator tends to $(-1)^k k!$, we obtain

$$\left(\frac{d}{dt}\frac{1}{\Gamma(t)}\right)_{t=-k} = (-1)^k k!.$$

If we substitute the value of the derivative $d[1/\Gamma(t)]/dt$ at integral values of t into the series (44), it follows for $\lambda = 1, 2, \cdots$ that

$$\begin{aligned}\pi N_\lambda(z) &= \frac{\partial J_\lambda(z)}{\partial\lambda} - (-1)^\lambda \frac{\partial J_{-\lambda}(z)}{\partial\lambda} \\ &= 2J_\lambda(z)\left(\log\frac{z}{2} + C\right) - \left(\frac{z}{2}\right)^{-\lambda}\sum_{n=0}^{\lambda-1}\frac{(\lambda-n-1)!}{n!}\left(\frac{z}{2}\right)^{2n} \\ &\quad - \left(\frac{z}{2}\right)^\lambda \frac{1}{\lambda!}\left\{\frac{1}{\lambda} + \frac{1}{\lambda-1} + \cdots + 1\right\} \qquad (45) \\ &\quad - \left(\frac{z}{2}\right)^\lambda \sum_{n=1}^{\infty}\frac{(-1)^n\left(\frac{z}{2}\right)^{2n}}{n!(n+\lambda)!}\cdot \\ &\quad \left\{\frac{1}{n+\lambda} + \frac{1}{n+\lambda-1} + \cdots + 1 + \frac{1}{n} + \frac{1}{n-1} + \cdots + 1\right\}\end{aligned}$$

and for $\lambda = 0$ that

$$\pi N_0(z) = 2J_0(z)\left(\log\frac{z}{2} + C\right) - 2\sum_{n=1}^{\infty}\frac{(-1)^n}{(n!)^2}\left(\frac{z}{2}\right)^{2n}\left\{\frac{1}{n} + \frac{1}{n-1} + \cdots + 1\right\}.$$

From the last expansion we get an idea of the *singularities* which can occur in connection with the solutions of Bessel's equation.

With the exception of the point $z = \infty$, which is an essential singularity for all nonidentically vanishing solutions, the origin is the only point at which the solutions of Bessel's equation can be singular. If λ is not an integer, the most general solution can be represented with the aid of the functions $J_\lambda(z)$ and $J_{-\lambda}(z)$ and thus at the origin can exhibit singularities only of the form z^λ and $z^{-\lambda}$, respectively. If $\lambda = n$ is an integer, besides a pole of order n at the origin the solutions can have only a logarithmic singularity of the form $z^n \log z$. For, every solution may be expressed as a linear combination of the functions $J_n(z)$ and $N_n(z)$, and these have no other singularities.

The Bessel functions $J_n(z)$ with integral index n, in particular, are just those solutions which remain regular at the origin.

§3. *Legendre Functions*

The Legendre functions and the higher spherical functions obtained from them by differentiation have already[1] been investigated as functions of a real variable and many of their properties have been derived. Now we shall develop integral representations for these functions by going over to complex variables $z = x + iy$; at the same time we shall try to find the remaining solutions of Legendre's equation. In the process it will become clear that the parameter n in the Legendre function $P_n(z)$ need not be restricted to positive integral values.[2]

1. Schläfli's Integral. From the representation (Ch. II, §8)

$$P_n(z) = \frac{1}{2^n n!}\frac{d^n}{dz^n}(z^2 - 1)^n$$

[1] In Ch. II, §8 and Ch. V, §10, 2.

[2] In connection with this paragraph, see in particular E. T. Whittaker and G. N. Watson. A Course of Modern Analysis, pp. 302–336, 4th ed., Cambridge University Press, Cambridge, 1927.

of the nth Legendre polynomial, one immediately obtains the expression

$$P_n(z) = \frac{1}{2\pi i}\int_C \frac{(\zeta^2 - 1)^n}{2^n(\zeta - z)^{n+1}}\,d\zeta \tag{46}$$

for arbitrary complex z, using Cauchy's integral formula. Here the path of integration C in the plane of the complex variable $\zeta = \xi + i\eta$ is taken in the positive sense about the point $\zeta = z$. From this expression, which is due to Schläfli, important consequences and generalizations can be derived. First we observe that Legendre's equation

$$\frac{d}{dz}\left((1 - z^2)\frac{dP_n}{dz}\right) + n(n + 1)P_n = 0$$

is a direct consequence of the integral representation (46). Indeed, differentiation under the integral sign gives the expression

$$\begin{aligned}
&\frac{n + 1}{2\pi i 2^n}\int_C \frac{(\zeta^2 - 1)^n}{(\zeta - z)^{n+3}}\left((n + 2)(1 - z^2) - 2z(\zeta - z) + n(\zeta - z)^2\right)d\zeta \\
&= \frac{n + 1}{2\pi i 2^n}\int_C \frac{(\zeta^2 - 1)^n}{(\zeta - z)^{n+3}}\left(2(n + 1)\zeta(\zeta - z) - (n + 2)(\zeta^2 - 1)\right)d\zeta \\
&= \frac{n + 1}{2\pi i 2^n}\int_C \frac{d}{d\zeta}\left(\frac{(\zeta^2 - 1)^{n+1}}{(\zeta - z)^{n+2}}\right)d\zeta
\end{aligned}$$

for the left side of the differential equation. Because the path of integration is closed and $(\zeta^2 - 1)^{n+1}/(\zeta - z)^{n+2}$ is single-valued, this expression vanishes identically in ζ. We can use this direct verification of Legendre's equation in order to extend the definition of $P_n(z)$ to arbitrary values of n which are not positive integers. Evidently Schläfli's integral (46) must represent a solution of Legendre's equation for arbitrary n whenever $(\zeta^2 - 1)^{n+1}/(\zeta - z)^{n+2}$ returns to its starting value as the path of integration is traversed once; for example, when the path is closed on the Riemann surface of the integrand. In this case, however, the function $P_n(z)$ will not in general be an integral rational—not even a single-valued analytic—function of z. Paths of this kind can be obtained in the following way: We cut the ζ-plane along the real axis from -1 to $-\infty$ and along an arbitrary path from the point 1 to the point z; the z-plane is cut analogously. For C we choose a closed curve, oriented in the

positive sense, which encloses the points $\zeta = z$ and $\zeta = +1$ but excludes the point $\zeta = -1$. The function

$$P_\nu(z) = \frac{1}{2\pi i}\int_C \frac{(\zeta^2 - 1)^\nu}{2^\nu(\zeta - z)^{\nu+1}}\,d\zeta \tag{47}$$

so defined, which is single-valued in the cut z-plane, is also called the *Legendre function of index* ν. It satisfies Legendre's differential equation

$$((1 - z^2)u')' + \nu(\nu + 1)u = 0 \tag{48}$$

and may be characterized as that uniquely determined solution which for $z = 1$ is finite and has the value[1]

$$P_\nu(1) = 1.$$

This property immediately becomes evident from the integral representation when z is allowed to approach 1. Since the above differential equation is transformed into itself by the substitution of $-\nu - 1$ for ν, we have the identity

$$P_\nu(z) = P_{-\nu-1}(z),$$

which is not easy to verify by calculation.

As can be seen from the representation, $P_\nu(z)$ satisfies the recursion formulas

$$\begin{aligned} P'_{\nu+1}(z) - zP'_\nu(z) &= (\nu + 1)P_\nu(z), \\ (\nu + 1)P_{\nu+1}(z) - z(2\nu + 1)P_\nu(z) + \nu P_{\nu-1}(z) &= 0, \end{aligned} \tag{49}$$

the second of which was derived in Ch. II, §8, 3 for integral ν.

2. Integral Representations of Laplace. If the real part of z is positive and $z \neq 1$, as we shall now assume, we can take for C the circle of radius $|\sqrt{z^2 - 1}|$ about the point z. This circle has the properties required above, as we see from the inequality $|z - 1|^2 < |z + 1||z - 1|$ which holds for $\mathfrak{R}_e(z) > 0, z \neq 1$. If we introduce the real variable of integration φ by setting $\zeta = z + \sqrt{z^2 - 1}\,e^{i\varphi}$, $|\varphi| \leq \pi$, Schläfli's integral yields at once the first integral repre-

[1] In fact, the second integral Q_ν, which will be defined in subsection 3, becomes logarithmically infinite at $z = 1$. Hence the same is true of any integral which is linearly independent of P_ν.

sentation of Laplace, valid also for $z = 1$:

$$P_\nu(z) = \frac{1}{\pi}\int_0^\pi (z + \sqrt{z^2 - 1}\cos\varphi)^\nu \, d\varphi, \tag{50}$$

in which the multiple-valued function $(z + \sqrt{z^2 - 1}\cos\varphi)^\nu$ is to be determined in such a way that for $\varphi = \pi/2$ it is equal to z^ν; here z^ν denotes the principal value of z^ν, in particular, a real value for positive z and real ν.

The formula $P_\nu = P_{-\nu-1}$ immediately yields the second Laplace representation

$$P_\nu(z) = \frac{1}{\pi}\int_0^\pi \frac{d\varphi}{(z + \sqrt{z^2 - 1}\cos\varphi)^{\nu+1}}. \tag{51}$$

It should be noted that for those z for which the expression $z + \sqrt{z^2 - 1}\cos\varphi$ vanishes on the path of integration the first representation fails to be valid only for $\nu \leq -1$, and the second only for $\nu \geq 0$. Therefore, at least one representation defines a single-valued branch of P_ν in the plane (except on a cut along the real axis from -1 to ∞.)

3. Legendre Functions of the Second Kind. The differential equation (48) must also have a second solution which is linearly independent of $P_\nu(z)$. This solution can be obtained easily from Schläfli's integral by taking a path of integration different from that used above. Such a path is given by the path $\mathfrak{A}$ in the form of a figure eight (see Figure 14, page 480), provided this path does not enclose the point z. Legendre's equation is again satisfied by the analytic function $Q_\nu(z)$ which is defined by the integral

$$Q_\nu(z) = \frac{-1}{4i\sin\nu\pi}\int_{\mathfrak{A}} \frac{1}{2^\nu}\frac{(\zeta^2 - 1)^\nu}{(z - \zeta)^{\nu+1}}\, d\zeta \tag{52}$$

where $\arg(\zeta^2 - 1) = 0$ on the intersection of $\mathfrak{A}$ with the positive real axis at the right of $\zeta = 1$. It is called a *Legendre function of the second kind*, and is regular and single-valued in the z-plane which has been cut along the real axis from $+1$ to $-\infty$. At first we explicitly assume that in this representation ν is not an integer, since otherwise the normalizing factor $1/\sin\nu\pi$ would become infinite. In case the real part of $\nu + 1$ is positive, we can contract the path of

integration and write (compare the calculation on page 483)

$$Q_\nu(z) = \frac{1}{2^{\nu+1}} \int_{-1}^{1} \frac{(1-\zeta^2)^\nu}{(z-\zeta)^{\nu+1}}\, d\zeta. \tag{53}$$

This formula is now applicable for non-negative integral ν also.

From representation (52) it is easily seen that Q_ν becomes logarithmically infinite at $z = 1$ and $z = -1$, since the path must pass through the lines joining z to the points $+1$ and -1.

An integral representation analogous to the Laplace integrals for $P_\nu(z)$ holds also for the functions $Q_\nu(z)$. In the above integral (53) we set

$$\zeta = \frac{e^\varphi \sqrt{z+1} - \sqrt{z-1}}{e^\varphi \sqrt{z+1} + \sqrt{z-1}}$$

and consider first real $z > 1$. After some calculation, we obtain

$$Q_\nu(z) = \int_0^\infty \frac{d\varphi}{(z + \sqrt{z^2-1}\cosh\varphi)^{\nu+1}} \qquad (\nu > -1), \tag{54}$$

in which the value of the integrand is determined in the same way as above.

4. Associated Legendre Functions. (Legendre functions of higher order.) For the Legendre functions of higher order, which we define by means of the equations

$$P_{\nu,h}(z) = (1-z^2)^{h/2} \frac{d^h}{dz^h} P_\nu(z),$$

$$Q_{\nu,h}(z) = (1-z^2)^{h/2} \frac{d^h}{dz^h} Q_\nu(z)$$

(see Ch. V, §10, 2, p. 327), we likewise obtain integral formulas by differentiating Schläfli's integral representation (47) and then substituting $\zeta = z + \sqrt{z^2-1}\, e^{i\varphi}$ (see subsection 2). Explicitly we write

$$P_{\nu,h}(z) = (i)^h \frac{(\nu+1)(\nu+2)\cdots(\nu+h)}{\pi} \int_0^\pi (z + \sqrt{z^2-1}\cos\varphi)^\nu \cos h\varphi\, d\varphi. \tag{55}$$

From this expression we see immediately, for example, that all the Legendre functions $P_{\nu,h}(z)$ with $h > 0$ vanish for $z = 1$.

§4. *Application of the Method of Integral Transformation to Legendre, Tchebycheff, Hermite, and Laguerre Equations*

The theory of the Legendre equation, as well as that of the orthogonal functions discussed in Chapter II, can be developed with the methods of the integral transformation described in §1. In the following we shall sketch briefly how this is done.

1. Legendre Functions. In the Legendre equation

$$L[u] = (1 - z^2)u'' - 2zu' = -\lambda(\lambda + 1)u \tag{56}$$

we are led, by means of the transformation

$$u(z) = \int_C K(z, \zeta)v(\zeta)\, d\zeta,$$

to the condition

$$\int_C \{(1 - z^2)K_{zz} - 2zK_z + \lambda(\lambda + 1)K\}v(\zeta)\, d\zeta = 0.$$

If we require the transformation kernel to satisfy the differential equation

$$(1 - z^2)K_{zz} - 2zK_z + \zeta(\zeta K)_{\zeta\zeta} = 0, \tag{57}$$

of which the function $K = 1/\sqrt{1 - 2z\zeta + \zeta^2}$ is a solution, and replace $L[K]$ by $-\zeta(\zeta K)_{\zeta\zeta}$ in the integral, we can transform the resulting integral using integration by parts; we obtain for $v(\zeta)$ the differential equation

$$\zeta(v\zeta)'' - \lambda(\lambda + 1)v = 0,$$

which has the solutions $v = \zeta^{\lambda}$ and $v = \zeta^{-\lambda-1}$. This leads to the integrals

$$\begin{aligned} P_\lambda(z) &= \frac{1}{2\pi i}\int_{C_1} \frac{\zeta^{-\lambda-1}}{\sqrt{1 - 2z\zeta + \zeta^2}}\, d\zeta, \\ Q_\lambda(z) &= \frac{1}{4i \sin \pi\lambda}\int_{C_2} \frac{\zeta^{-\lambda-1}}{\sqrt{1 - 2z\zeta + \zeta^2}}\, d\zeta \end{aligned} \tag{58}$$

where C_1 and C_2 (see Figures 16 and 17) denote the curves on the Riemann surface of the integrand.

By means of the transformation

$$\zeta = z + \sqrt{z^2 - 1}\cos\varphi$$

and a suitable deformation of the path of integration, we at once obtain the Laplace integrals (51) and (54)

$$P_\lambda(z) = \frac{1}{\pi}\int_0^\pi (z + \sqrt{z^2-1}\cos\varphi)^{-\lambda-1}\,d\varphi,$$

$$Q_\lambda(z) = \int_0^\infty (z + \sqrt{z^2-1}\cosh\varphi)^{-\lambda-1}\,d\varphi \quad (\lambda < -1).$$

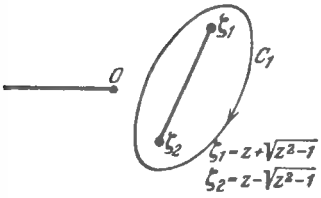


Figure 16

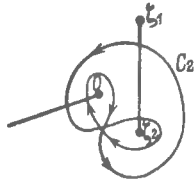


Figure 17

The kernel selected above,

$$K(z,\zeta) = \frac{1}{\sqrt{1-2z\zeta+\zeta^2}},$$

as well as any other which satisfies equation (57), is a generating function of the Legendre equation. For, the coefficient $u_n(z)$ in the series expansion

$$K(z,\zeta) = \sum_0^\infty u_n(z)\zeta^n$$

of such a kernel is an integral of the above form,

$$u_n(z) = \frac{1}{2\pi i}\oint \frac{K(z,\zeta)}{\zeta^{n+1}}\,d\zeta;$$

since the path of integration is closed, $u_n(z)$ is a solution of equation (56) for $\lambda = n$.

2. Tchebycheff Functions. In the case of Tchebycheff's equation

$$L[u] = (1-z^2)u'' - zu' = -\lambda^2 u, \tag{59}$$

we take K to be a solution of the differential equation

$$(1-z^2)K_{zz} - zK_z + \zeta(\zeta K_\zeta)_\zeta = 0; \tag{60}$$

for example, we take $K(z,\zeta) = 1 - \zeta^2/(1-2z\zeta+\zeta^2)$, which leads to

solutions of the form

$$P_\lambda(z) = \frac{1}{2\pi i}\int_{C_1} \frac{1-\zeta^2}{1-2z\zeta+\zeta^2}\,\zeta^{-\lambda-1}\,d\zeta,$$
$$Q_\lambda(z) = \frac{1}{2\pi i}\int_{C_2} \frac{1-\zeta^2}{1-2z\zeta+\zeta^2}\,\zeta^{-\lambda-1}\,d\zeta. \tag{61}$$

Here C_1 and C_2 are closed curves on the Riemann surface of the integrand and enclose the zeros of the denominator

$$\zeta_1 = z + \sqrt{z^2-1}; \qquad \zeta_2 = z - \sqrt{z^2-1}$$

(see Figure 18).

Application of Cauchy's integral theorem yields

$$P_\lambda(z) = (z+\sqrt{z^2-1})^\lambda,$$
$$Q_\lambda(z) = (z-\sqrt{z^2-1})^\lambda. \tag{62}$$

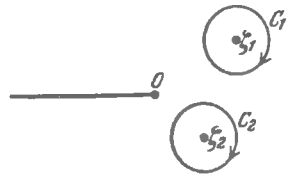


Figure 18

The sum

$$T_\lambda(z) = \frac{1}{2^\lambda}(P_\lambda(z)+Q_\lambda(z)) = \frac{(z+\sqrt{z^2-1})^\lambda + (z-\sqrt{z^2-1})^\lambda}{2^\lambda},$$

which is represented also by the integral

$$T_\lambda(z) = \frac{1}{2^\lambda}\frac{1}{2\pi i}\int_C \frac{1-\zeta^2}{1-2z\zeta+\zeta^2}\,\zeta^{-\lambda-1}\,d\zeta$$

in which C now encloses both points ζ_1 and ζ_2, goes over into the n-th Tchebycheff polynomial for $\lambda = n$.

3. Hermite Functions. In the case of the Hermite equation

$$L[u] = u'' - 2zu' = -2\lambda u \tag{63}$$

we require K to satisfy equation

$$K_{zz} - 2zK_z + 2\zeta K_\zeta = 0, \tag{64}$$

which has the function $e^{2z\zeta-\zeta^2}$ as a solution. If for C we take one of the curves represented in Figure 19, we obtain the solutions

$$P_\lambda(z) = \frac{1}{\pi i}\int_{C_1} \frac{e^{-\zeta^2+2z\zeta}}{\zeta^{\lambda+1}}\,d\zeta,$$
$$Q_\lambda(z) = \frac{1}{\pi i}\int_{C_2} \frac{e^{-\zeta^2+2z\zeta}}{\zeta^{\lambda+1}}\,d\zeta. \tag{65}$$

Their arithmetic mean,

$$H_\lambda(z) = \tfrac{1}{2}(P_\lambda(z) + Q_\lambda(z)),$$

in other words, the integral

$$H_\lambda(z) = \frac{1}{2\pi i}\int_C \frac{e^{-\zeta^2+2z\zeta}}{\zeta^{\lambda+1}}\,d\zeta,$$

where C is the loop in Figure 21, becomes the Hermite polynomial $H_n(z)$ for $\lambda = n$.

If $\mathcal{R}_e\,(\lambda) < 0$, we can contract the path of integration to the origin and obtain as solutions—up to factors independent of z—the integrals

$$\int_0^\infty \frac{e^{-\zeta^2+2z\zeta}}{\zeta^{\lambda+1}}\,d\zeta \tag{66}$$

and

$$\int_{-\infty}^\infty \frac{e^{-\zeta^2+2z\zeta}}{\zeta^{\lambda+1}}\,d\zeta. \tag{67}$$

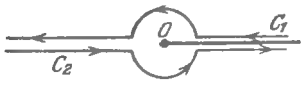


Figure 19

4. Laguerre Functions. Correspondingly, in the Laguerre equation

$$L[u] = zu'' + (1 - z)u' = -\lambda u \tag{68}$$

we impose on $K(z, \zeta)$ the partial differential equation

$$zK_{zz} + (1 - z)K_z + \zeta K_\zeta = 0 \tag{69}$$

and arrive at integrals of the form

$$\int_C \frac{e^{-z\zeta/(1-\zeta)}}{1-\zeta}\zeta^{-\lambda-1}\,d\zeta. \tag{70}$$

In the choice of the path of integration C, it must be noted that the integrand has an essential singularity at the point $\zeta = 1$. In particular if we take for C the curve shown in Figure 20, the integral

$$L_\lambda(z) = \frac{1}{2\pi i}\int_C \frac{e^{-z\zeta/(1-\zeta)}}{1-\zeta}\zeta^{-\lambda-1}\,d\zeta \tag{71}$$

represents solutions which are essentially identical with the Laguerre polynomials in the case $\lambda = n$.

By means of the transformation

$$u = \frac{\zeta}{1-\zeta},$$

(71) acquires the form

$$L_\lambda(z) = \frac{1}{2\pi i}\int_C \frac{e^{-uz}}{u^{\lambda+1}}(1+u)^\lambda\, du; \tag{72}$$

now C denotes the path in Figure 21.

Furthermore, we note that, as in the case of Legendre's equation, every solution of the partial differential equation associated with each equation discussed here can be considered as a *generating function* of a family of solutions of the original differential equation. In

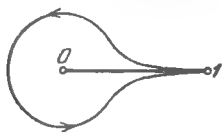


Figure 20

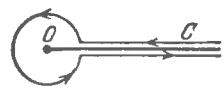


Figure 21

particular, the special kernels we have used define, by their power series expansions, the Tchebycheff, Hermite, and Laguerre polynomials.

§5. *Laplace Spherical Harmonics*

The Laplace spherical harmonics $Y_n(\theta, \varphi)$ were introduced in Ch. V, §8, p. 314 as the everywhere-regular eigenfunctions, corresponding to the eigenvalues $\lambda = n(n+1)$, of the differential equation

$$\Delta^* Y + \lambda Y = \frac{1}{\sin^2\theta} Y_{\varphi\varphi} + \frac{1}{\sin\theta}(\sin\theta Y_\theta)_\theta + \lambda Y = 0 \tag{73}$$

on the sphere. We see that the functions $r^n Y_n = U_n$ are then polynomials in the rectangular coordinates x, y, z which satisfy the differential equation $\Delta U = 0$ and are homogeneous of degree n. Conversely, every integral rational solution U_n of the differential equation $\Delta U = 0$ which is homogeneous of degree n becomes a Laplace spherical harmonic when divided by r^n. Since a polynomial which is homogeneous of degree n has $(n+1)(n+2)/2$ coefficients, and since the condition $\Delta U_n \equiv 0$ determines $(n-1)n/2$ linear homogeneous relations between these coefficients—for, ΔU_n is homogeneous of $(n-2)$-nd degree—, the U_n have at least

$$\frac{(n+1)(n+2)}{2} - \frac{(n-1)n}{2} = 2n+1$$

independent coefficients; therefore there must be at least $2n + 1$ linearly independent spherical harmonics of n-th order.

In this paragraph it will be shown that the aforementioned conditions are mutually independent, so that there exist precisely $2n + 1$ linearly independent spherical harmonics of n-th order. It will be shown also that these functions Y_n really represent all the eigenfunctions of our eigenvalue problem, and hence that the values $\lambda = n(n + 1)$ represent all the eigenvalues. Finally, these functions will be expressed explicitly in terms of associated Legendre functions, with which we are familiar from §3 and Ch. V, §10, 2, p. 327. We begin with the last point.

1. Determination of $2n + 1$ Spherical Harmonics of n-th Order. We again obtain special spherical harmonics by means of the familiar trial substitution $Y(\theta, \varphi) = p(\varphi)q(\theta)$. We make this substitution in equation (73) for $\lambda = n(n + 1)$, and denote differentiation with respect to φ by a prime, and differentiation with respect to θ by a dot; equation (73) then becomes

$$\frac{p''(\varphi)}{p(\varphi)} = -\frac{(\sin\theta\dot{q})^{\cdot}\sin\theta}{q} - n(n+1)\sin^2\theta = -\rho,$$

in which ρ must be a constant. Thus we obtain for q the equation

$$(\sin\theta\dot{q})^{\cdot} + \left(n(n+1)\sin\theta - \frac{\rho}{\sin\theta}\right)q = 0,$$

in which the parameter ρ must be determined in such a way that a solution exists which is regular for $\theta = 0$ and $\theta = \pi$. Substituting $z = \cos\theta$, and denoting differentiation with respect to z by a prime, this equation immediately becomes

$$((1 - z^2)q')' + \left(n(n+1) - \frac{\rho}{1 - z^2}\right)q = 0$$

with the boundary condition of regularity for $z = +1$ and $z = -1$; this problem is familiar to us, in a somewhat different form, from Ch. V, §10, 2, p. 327. We already know the solutions $\rho = h^2$ and $q = P_{n,h}(z)$, where $P_{n,h}(z)$ is the Legendre function of h-th order, and $P_{n,h} = (1 - z^2)^{h/2}(d^h/dz^h)P_n(z)$; h can assume the values $0, 1, 2, \cdots, n$. Since $p''(\varphi) + h^2p(\varphi) = 0$, p is given as $a_h \cos h\varphi + b_h \sin h\varphi$; since $Y = pq$, we at once obtain

$$Y(\theta, \varphi) = (a_h \cos h\varphi + b_h \sin h\varphi)P_{n,h}(\cos\theta)$$

as a solution of (73). Therefore the function

$$(74)\quad Y_n = \frac{a_{n,0}}{2} P_n(\cos\theta) + \sum_{h=1}^{n} (a_{n,h}\cos h\varphi + b_{n,h}\sin h\varphi) P_{n,h}(\cos\theta)$$

is a spherical harmonic of dimension n depending upon $2n + 1$ arbitrary linear parameters. It will soon be apparent that this function is the most general possible. The functions $\cos h\varphi\, P_{n,h}(\cos\theta)$, $\sin h\varphi\, P_{n,h}(\cos\theta)$ are all linearly independent, since they are mutually orthogonal; they will be called symmetric spherical harmonics of order n.

2. Completeness of the System of Functions. From earlier theorems we see at once that the $2n + 1$ functions $\cos h\varphi\, P_{n,h}(\cos\theta)$, $\sin h\varphi\, P_{n,h}(\cos\theta)$ form a complete orthogonal system of functions on the sphere. The functions $\sin h\varphi$, $\cos h\varphi$ form a complete system in the variable φ, and the functions $P_{n,h}(z)$ for every h form a complete system in z because the system of all eigenfunctions of an eigenvalue problem is always complete (see Ch. VI, §3, 1). In order to prove that the system of functions is complete, we need only recall the theorem of Ch. II, §1, 6, which contains a general rule for constructing a complete system of functions in two variables from two complete systems in one variable.

From this result it follows immediately that equation (73) has no eigenfunctions other than those indicated, and thus can have no eigenvalues except the values $n(n + 1)$. All the questions which were raised above have thus been answered. (Note that this constitutes a transcendental proof of the algebraic fact that there exist precisely $2n + 1$ linearly independent functions Y_n.)

This fact may, of course, be proved easily from the standpoint of algebra alone. Consider any polynomial u of x, y, z which is homogeneous of degree n: $u = \sum a_{rst} x^r y^s z^t \quad (r + s + t = n)$. Then every coefficient a_{rst} may be represented, up to a constant factor, as a derivative of the form $\partial^n u/\partial x^r \partial y^s \partial z^t$. If $\Delta u = 0$, every derivative of the form $\partial u^m/\partial x^\alpha \partial y^\beta \partial z^\gamma$ can be written in such a way that differentiation with respect to x occurs no more than once; we use the differential equation $u_{xx} = -u_{yy} - u_{zz}$ to eliminate all derivatives of u with respect to x of order higher than one (for example, $\partial^3 u/\partial x^2 \partial y = -\partial^3 u/\partial y^3 - \partial^3 u/\partial z^2 \partial y$). Thus, if $\Delta u = 0$, all coefficients of u are linear combinations of the $2n + 1$ coefficients $a_{0,0,n}, a_{0,1,n-1}, \cdots, a_{0,n,0}; a_{1,0,n-1}, a_{1,1,n-2}, \cdots, a_{1,n-1,0}$, which may be chosen arbitrarily.

3. Expansion Theorem. Since the functions (74) represent all the eigenfunctions of our problem, our earlier theorems (see e.g. Ch. V, §14, 5) imply that every function $g(\theta, \varphi)$ which is continuous together with its derivatives up to the second order on the sphere may be expanded in an absolutely and uniformly convergent series in terms of the spherical harmonics

$$g(\theta, \varphi) = \sum_{n=0}^{\infty} \left[a_{n,0} P_n(\cos\theta) + \sum_{h=1}^{n} (a_{n,h} \cos h\varphi + b_{n,h} \sin h\varphi) P_{n,h}(\cos\theta) \right]$$

where, in view of the formulas of Ch. V, §10, p. 327, the coefficients $a_{n,0}$, $a_{n,h}$, $b_{n,h}$ are determined from the relations

$$a_{n,0} = \frac{2n+1}{4\pi} \int_{-\pi}^{\pi} \int_{0}^{\pi} g(\theta, \varphi) P_n(\cos\theta) \sin\theta \, d\theta \, d\varphi,$$

$$(75) \quad a_{n,h} = \frac{2n+1}{2\pi} \frac{(n-h)!}{(n+h)!} \int_{-\pi}^{\pi} \int_{0}^{\pi} g(\theta, \varphi) P_{n,h}(\cos\theta) \cos h\varphi \sin\theta \, d\theta \, d\varphi,$$

$$b_{n,h} = \frac{2n+1}{2\pi} \frac{(n-h)!}{(n+h)!} \int_{-\pi}^{\pi} \int_{0}^{\pi} g(\theta, \varphi) P_{n,h}(\cos\theta) \sin h\varphi \sin\theta \, d\theta \, d\varphi.$$

Extension of this result to more general functions $g(\theta, \varphi)$ need not concern us here.

4. The Poisson Integral. We can now write the solution of the boundary value problem of potential theory for a sphere of radius 1 with boundary values $g(\theta, \varphi)$ explicitly in the form

$$u = \sum_{n=0}^{\infty} r^n \left[a_{n,0} P_n(\cos\theta) + \sum_{h=1}^{n} (a_{n,h} \cos h\varphi + b_{n,h} \sin h\varphi) P_{n,h}(\cos\theta) \right].$$

If we introduce the integral representations (75), we can interchange summation and integration since the convergence is uniform for $r \leq r_0 < 1$. The summation may then be carried out in closed form; in doing this it is easiest to assume at first $\theta = 0, \varphi = 0$, and then to note that the result must hold for an arbitrary point θ, φ, since any point on the sphere may be chosen as the north pole.

Since $P_n(1) = 1$, $P_{n,h}(1) = 0$ $(h = 1, 2, \cdots, n)$, we obtain

$$4\pi u(r, 0, 0) = \int_{-\pi}^{\pi} \int_{0}^{\pi} \left\{ \sum_{n=0}^{\infty} (2n+1) r^n P_n(\cos\theta) \right\} g(\theta, \varphi) \sin\theta \, d\theta \, d\varphi;$$

here the sum may be obtained in closed form by using the relation

$$\sum_{n=0}^{\infty} (2n+1) h^n P_n(z) = \frac{1 - h^2}{(1 - 2hz + h^2)^{3/2}},$$

which can be easily derived from the equation of definition

$$\sum_{n=0}^{\infty} h^n P_n(z) = (1 - 2hz + h^2)^{-1/2}$$

with the aid of the recursion formula (49). After carrying out the summation and again thinking of the pole of the sphere as displaced, we can write the result with complete generality in the form

$$4\pi u(r\ \theta,\ \varphi,) =$$

$$(76) \qquad (1 - r^2) \int_{\pi}^{\pi}\int_{0}^{\pi} \frac{g(\theta', \varphi') \sin\theta'\, d\theta'\, d\varphi'}{\{r^2 - 2r[\cos\theta \cos\theta' + \sin\theta \sin\theta' \cos(\varphi - \varphi')] + 1\}^{3/2}}.$$

This so-called *Poisson integral* expresses the potential function at interior points in terms of its boundary values; it no longer contains any explicit reference to the spherical harmonics. In Volume II we shall return to this integral in the framework of potential theory.

5. The Maxwell-Sylvester Representation of Spherical Harmonics. An entirely different representation for the spherical harmonics, connected with the physical meaning of a potential, was given by Maxwell.[1] In this section we shall investigate the properties of spherical functions in connection with Maxwell's basic ideas and a supplementary remark of Sylvester, and thus arrive at a new development of our theory.

We begin with the potential $1/r = 1/\sqrt{x^2 + y^2 + z^2}$ corresponding to a unit mass concentrated at the origin, and note that every derivative $v = \partial^n u/\partial x^\alpha \partial y^\beta \partial z^\gamma$ $(n = \alpha + \beta + \gamma)$ of a potential function u is again a solution of the potential equation $\Delta v = 0$, since from $\Delta u = 0$ we obtain by differentiation

$$0 = \frac{\partial}{\partial x} \Delta u = \Delta \frac{\partial u}{\partial x},$$

etc. Hence, if a, b, c are constants, the function

$$a \frac{\partial \frac{1}{r}}{\partial x} + b \frac{\partial \frac{1}{r}}{\partial y} + c \frac{\partial \frac{1}{r}}{\partial z}$$

[1] A Treatise on Electricity and Magnetism, Vol. 1, pp. 179–214, 2nd ed., Clarendon Press, Oxford, 1881.

is also a potential function. Using the symbolic linear form

$$L = a\frac{\partial}{\partial x} + b\frac{\partial}{\partial y} + c\frac{\partial}{\partial z},$$

we can write it in the form $L\frac{1}{r}$ or in the form

$$\alpha\frac{\partial\frac{1}{r}}{\partial\nu},$$

where $\alpha = \sqrt{a^2 + b^2 + c^2}$ and $\partial/\partial\nu$ stands for differentiation in that direction ν whose direction cosines are proportional to a, b, c.[1] Physically, this potential corresponds to a dipole of moment α and direction ν. More generally, in the expression

$$u = C\frac{\partial^n\frac{1}{r}}{\partial\nu_1\partial\nu_2\cdots\partial\nu_n} = CL_1L_2\cdots L_n\frac{1}{r} \tag{77}$$

we obtain a potential which corresponds to a "*multipole*" with the axes $\nu_1, \nu_2, \cdots, \nu_n$. The L_i denote linear forms in the operators $\partial/\partial x$, $\partial/\partial y$, $\partial/\partial z$, and their coefficients a_i, b_i, c_i define the axis directions ν_i. It is easily seen that the potential u has the form

$$u = U_n(x, y, z)r^{-2n-1}, \tag{78}$$

where U_n is a homogeneous polynomial of degree n in x, y, z. The function U_n itself satisfies the potential equation $\Delta U_n = 0$, as is seen from the following general theorem: *If $u(x, y, z)$ is a solution of the potential equation, then $(1/r)u(x/r^2, y/r^2, z/r^2)$ is also a solution.*[2] For $r = 1$, such functions $U_n(x, y, z)$ according to our earlier definitions(Ch. V, §9, 1) are spherical harmonics of order n.

Since each of the n directions occurring in (77) is determined by two parameters, and an additional arbitrary constant factor appears in the potential u, there are $2n + 1$ arbitrary constants in all. It may be conjectured, therefore, that all the spherical harmonics of order n can in fact be represented in the form (77). We shall prove

[1] If complex values are admitted for the a, b, c, necessary precautions must naturally be taken in the case of those triples for which $a^2 + b^2 + c^2 = 0$.

[2] The proof of this theorem is obtained immediately from the polar coordinate form of the potential equation (see Ch. IV, §8, 2).

this rigorously by first representing the $2n + 1$ linearly independent symmetric spherical harmonics $P_{n,h}(\cos\theta)\sin h\varphi$, $P_{n,h}(\cos\theta)\cos h\varphi$ in terms of multipole potentials. Subsequently it will follow that every n-th order spherical harmonic is given by a sum of multipole potentials. Finally, we shall show that every sum of several such multipole potentials is equal to the potential of a single multipole, which can be obtained by means of a simple geometric construction.

The $2n + 1$ symmetric spherical harmonics of subsection 1 are obtained easily by considering symmetric multipoles. Let n axes with the directions $\nu_1, \nu_2, \cdots, \nu_n$ be arranged symmetrically in the x, y-plane in such a way that any two consecutive axes form the angle $2\pi/n$ with each other. If we set

$$\frac{\partial^n \frac{1}{r}}{\partial\nu_1\partial\nu_2\cdots\partial\nu_n} = u_n = U_n r^{-2n-1} \tag{79}$$

and note that the left side is invariant with respect to rotations of the sphere about the z-axis through the angle $2\pi/n$, we see at once that the n-th order spherical harmonic

$$u_n r^{n+1} = U_n r^{-n} = Y_n(\theta, \varphi),$$

which certainly does not vanish identically,[1] has the period $2\pi/n$ when considered as a function of φ. By subsection 3 every spherical harmonic of order n permits the representation

$$\sum_{h=0}^{n} (a_{n,h}\cos h\varphi + b_{n,h}\sin h\varphi)P_{n,h}(\cos\theta).$$

It follows that $Y_n(\theta, \varphi)$ has the form

$$\begin{aligned} Y_n(\theta, \varphi) &= [a_{n,n}\cos n\varphi + b_{n,n}\sin n\varphi]\,P_{n,n}(\cos\theta) + a_{n,0}\,P_{n,0}(\cos\theta) \\ &= \alpha\cos n(\varphi - \varphi_0)P_{n,n}(\cos\theta) + \beta P_{n,0}(\cos\theta). \end{aligned} \tag{80}$$

The two terms in (80) are the only ones which are periodic in φ with period $2\pi/n$. The term $P_{n,0}(\cos\theta)$ can be expressed separately in terms of a derivative of r^{-1}. We have

$$\frac{\partial^n}{\partial z^n}\left(\frac{1}{r}\right) = \frac{(-1)^n n!}{r^{n+1}} P_n(\cos\theta). \tag{80'}$$

[1] That no multipole potential can vanish identically will be proved on page 519.

The fact that the right side in (80′) is a multiple of $r^{-n-1} P_{n,0}(\cos \theta)$ follows immediately from the remarks made above. The calculation of the constant factor can be carried out by induction, using the recursion formulas (49).

In order to obtain a multipole representation for the remaining spherical harmonics of n-th order, we note that because of (80) the potential u_n can be decomposed as follows:

$$u_n = f(x, y) g\left(\frac{z}{r}\right) r^{-n-1}$$

where $f(x, y) = \alpha \cos n(\varphi - \varphi_0)$, $f(0, 0) = 0$. We replace n by h in this expression and then differentiate $n - h$ times with respect to z. The resulting potential function $u_{n,h}$ again has the form

$$u_{n,h} = f(x, y) g\left(\frac{z}{r}\right) r^{-n-1},$$

from which we conclude that the n-th order spherical harmonic

$$Y_n(\theta, \varphi) = u_{n,h} r^{n+1}$$

must have the form $\alpha \cos h(\varphi - \varphi_0)\, \omega(\theta)$. Therefore by subsection 1 it necessarily has the form

$$\text{const.} \times \cos h(\varphi - \varphi_0) P_{n,h}(\cos \theta). \tag{81}$$

Conversely, since one of the axes can be chosen arbitrarily, every function of this family is obtained by our procedure.

Since according to subsection 2 every spherical harmonic of order n may be expressed as a sum of $2n + 1$ spherical harmonics of the form (81), it follows at once that we obtain every n-th order spherical harmonic by forming sums of multipole potentials

$$u = \sum_{i+k+l=n} a_{ikl} \frac{\partial^n \frac{1}{r}}{\partial x^i \partial y^k \partial z^l}. \tag{82}$$

That, conversely, every such sum is a spherical harmonic of order n is obvious according to page 510. In fact, any particular spherical harmonic is represented infinitely often when we let the coefficients a_{ikl} run through all possible values. We shall now elaborate this point.

We first prove that every sum of the above form is the potential

of a single multipole with suitable axes. To do this we make use of a symbolic notation, considering the homogeneous polynomial of n-th degree

$$H(\xi, \eta, \zeta) = \sum_{i+k+l=n} a_{ikl}\xi^i\eta^k\zeta^l$$

and writing our potential in the form H/r, in which we replace the variables ξ, η, ζ in H by the differentiation symbols $\partial/\partial x, \partial/\partial y, \partial/\partial z$. Since with this meaning of ξ, η, ζ the function $(\xi^2 + \eta^2 + \zeta^2)/r$ is identically zero, we have $H/r = H_1/r$ provided the difference $H - H_1$, as a polynomial in the indeterminates ξ, η, ζ, can be expressed in the form $Q \cdot (\xi^2 + \eta^2 + \zeta^2)$, where Q denotes a homogeneous polynomial of $(n - 2)$-nd degree in ξ, η, ζ.

At this point we need a simple theorem employed by Sylvester:[1] *Given any homogeneous polynomial of n-th degree $H(\xi, \eta, \zeta)$, we can determine n linear forms $L_1, L_2, \cdots, L_n$ and a polynomial $Q(\xi, \eta, \zeta)$ of degree $n - 2$ in such a way that a relation of the form*

$$H = C \cdot L_1L_2 \cdots L_n + Q \cdot (\xi^2 + \eta^2 + \zeta^2)$$

holds. If H is real, the linear forms $L_1, L_2, \cdots, L_n$ are determined uniquely, up to constant factors, by the condition that their coefficients be real. The proof of this theorem, as well as the geometric characterization of the forms L_i, will be postponed to the end of the section so as not to interrupt the train of thought. Our assertion concerning the representation of the potential (82) by a single multipole follows at once from Sylvester's theorem. For, if ν_i denotes the axial direction perpendicular to the plane $L_i = 0$, we obtain

$$u = H\frac{1}{r} = C\frac{\partial^n \frac{1}{r}}{\partial\nu_1\partial\nu_2 \cdots \partial\nu_n},$$

which furnishes the desired representation.

Our theory is now outlined in its essential points. We shall give our discussion a somewhat different turn which avoids dependence on the results of subsections 1 and 2 and puts more stress on the

[1] J. J. Sylvester, Note on Spherical Harmonics, Phil. Mag., Vol. 2m, 1876, pp. 291–307 and 400. Collected Mathematical Papers, Vol. 3, pp. 37–51, Cambridge University Press, Cambridge, 1909.

purely algebraic nature of our theorems, although at the same time it relaxes the connection with the explicit representations. First we remark that two functions H/r and H_1/r are identical if and only if the difference $H^*(\xi, \eta, \zeta) = H(\xi, \eta, \zeta) - H_1(\xi, \eta, \zeta)$ is divisible by $\xi^2 + \eta^2 + \zeta^2$. As already pointed out, the first part of the assertion is obvious. In order to prove the second part, we must show that a relation $H^*/r = 0$ implies that the homogeneous polynomial $H^*(\xi, \eta, \zeta)$ is divisible by $\xi^2 + \eta^2 + \zeta^2$.[1] Now by Sylvester's theorem

$$H^* = C \cdot L_1^* L_2^* \cdots L_n^* + Q^* \cdot (\xi^2 + \eta^2 + \zeta^2), \tag{83}$$

where $L_1^*, L_2^*, \cdots, L_n^*$ denote linear forms which, in the case of real H^*, may be supposed real. If one of the linear forms L_i^* vanishes identically, our assertion is clear. However, if none of the linear forms vanishes identically, we have

$$H^* \frac{1}{r} = C \cdot L_1^* L_2^* \cdots L_n^* \frac{1}{r} = C \frac{\partial^n \frac{1}{r}}{\partial \nu_1^* \partial \nu_2^* \cdots \partial \nu_n^*},$$

and the multipole potential on the right side, because of the singularity at the origin, can vanish in the whole space only if $C = 0$. Otherwise, for suitable m, $0 \leq m < n$, we would have

$$\frac{\partial^m \frac{1}{r}}{\partial \nu_1 \cdots \partial \nu_m} = v_m \neq 0, \qquad \frac{\partial v_m}{\partial \nu_{m+1}} = 0;$$

hence v_m would have to have constant values on every parallel to the axis ν_{m+1}, which is impossible because of the singularity at the origin. We thus have

$$H^*(\xi, \eta, \zeta) = Q^*(\xi, \eta, \zeta) \cdot (\xi^2 + \eta^2 + \zeta^2),$$

which is what we wished to prove.

Clearly, every homogeneous function $H(\xi, \eta, \zeta)$ of degree n can be expressed in one and only one way in the form

$$H(\xi, \eta, \zeta) = G_n(\eta, \zeta) + \xi G_{n-1}(\eta, \zeta) + (\xi^2 + \eta^2 + \zeta^2) \cdot Q(\xi, \eta, \zeta). \tag{84}$$

Here G_n denotes a homogeneous function of degree n in η, ζ alone, G_{n-1} a homogeneous function of degree $(n - 1)$, and Q a homogeneous

[1] See the work by A. Ostrowski referred to on p. 521, footnote 2.

function of degree $(n - 2)$. The difference between two functions $H(\xi, \eta, \zeta)$ and $\bar{H}(\xi, \eta, \zeta)$ of degree n is divisible by $\xi^2 + \eta^2 + \zeta^2$ if and only if the associated functions G_n, G_{n-1}, $\bar{G}_n$, $\bar{G}_{n-1}$ satisfy the relations

$$G_n = \bar{G}_n, \qquad G_{n-1} = \bar{G}_{n-1}$$

identically. Since we have $2n + 1$ coefficients at our disposal in the functions G_n and G_{n-1}, the lemma just proved shows that we have precisely $2n + 1$ linearly independent potentials in the form H/r. Hence we obtain every spherical harmonic of degree n as a sum of potentials of multipoles. In order actually to obtain the representation of the spherical functions in this form one must use, in addition to this pure existence proof, an argument analogous to that carried out above.

Finally we shall prove Sylvester's theorem using a simple idea from algebraic geometry. By Bezout's theorem, the n-th degree cone $H(\xi, \eta, \zeta) = 0$ in ξ, η, ζ-space intersects the absolute cone $\xi^2 + \eta^2 + \zeta^2 = 0$ in precisely $2n$ edges; all multiple intersections are correctly weighted. We connect the $2n$ edges by means of n planes given by the equations

$$L_i(\xi, \eta, \zeta) = a_i\xi + b_i\eta + c_i\zeta = 0 \quad (i = 1, 2, \cdots, n)$$

in such a way that each plane contains two of the edges and every edge is accounted for once. Multiple edges appear according to their multiplicity.[1] We now consider the bundle of cones of n-th order

$$\lambda H + \mu L_1 L_2 \cdots L_n = 0$$

[1] We can make the meaning of this rule more precise without referring to the difficult general algebraic elimination theory: We uniformize the form $\xi^2 + \eta^2 + \zeta^2 = 0$ by setting

$$(*) \qquad \xi = \frac{1 - t^2}{1 + t^2}, \qquad \eta = \frac{2t}{1 + t^2}, \qquad \zeta = i = \sqrt{-1}.$$

A homogeneous function $H(\xi, \eta, \zeta)$ of degree n is then transformed by (*) into a rational function $H^*(t)$ of degree $2n$, whose zeros determine the common edges of the cones $H(\xi, \eta, \zeta) = 0$ and $\xi^2 + \eta^2 + \zeta^2 = 0$. We shall say that a common edge of these cones counts k-fold if on it $H^*(t)$, as a function of t, has a precisely k-fold zero. The linear forms $L_1, L_2, \cdots, L_n$ are now to be chosen in such a way that every k-fold curve of intersection of the cone $H = 0$ with the absolute cone is also a k-fold edge of intersection of the family of planes $L_1, L_2, \cdots, L_n = 0$. That this rule may be realized in every case is easily seen.

containing the two parameters λ and μ. Every cone of this bundle intersects the absolute cone in the $2n$ given fixed edges. Selecting any one edge of the absolute conic not one of the above fixed $2n$ edges), we determine the ratio λ/μ in such a way that the n-th degree cone

$$\lambda H + \mu L_1 L_2 \cdots L_n = 0$$

passes also through this edge, which is certainly possible and which leads to a value for λ/μ different from zero and infinity. The new cone of degree n then has more than $2n$ lines of intersection in common with a cone of second degree, which is impossible unless it contains the cone of second degree completely. This case arises if and only if the left side of the equation contains the expression $\xi^2 + \eta^2 + \zeta^2$ as a factor,[1] i.e. if

$$\lambda H + \mu L_1 L_2 \cdots L_n = Q \cdot (\xi^2 + \eta^2 + \zeta^2).$$

The proof of Sylvester's theorem is now complete.[2] At the same time a simple geometric interpretation is given for the axes of the multipole associated with a spherical harmonic.

Concerning reality considerations, one must note that although all lines of intersection are imaginary for real H, they are pairwise complex conjugates; thus there is just one way of projecting them onto n real planes.

[1] The first part of the assertion is obvious. The second is proved most simply by writing the given form, in accordance with (84), as

$$G_n(\eta, \zeta) + \xi G_{n-1}(\eta, \zeta) + (\xi^2 + \eta^2 + \zeta^2)Q(\xi, \eta, \zeta).$$

Now if η, ζ is any pair of values for which $\eta^2 + \zeta^2 \neq 0$, the two equations

$$0 = G_n(\eta, \zeta) + \sqrt{-(\eta^2 + \zeta^2)}G_{n-1}(\eta, \zeta)$$

and

$$0 = G_n(\eta, \zeta) - \sqrt{-(\eta^2 + \zeta^2)}G_{n-1}(\eta, \zeta)$$

hold simultaneously. We conclude at once

$$G_n(\eta, \zeta) = G_{n-1}(\eta, \zeta) = 0.$$

Thus G_n and G_{n-1} vanish for all pairs of values η, ζ with $\eta^2 + \zeta^2 \neq 0$; hence they vanish identically in η and ζ.

[2] This algebraic theorem was used by Sylvester, loc. cit., without proof. The necessity of obtaining a proof for it was indicated by A. Ostrowski. See Ostrowski, Mathematische Miszellen, I, Die Maxwellsche Erzeugung der Kugelfunktionen, Deutsch. Math.-Ver. Jahresber., Vol. 33, 1925, pp. 245–251.

§6. *Asymptotic Expansions*

Asymptotic expressions for our functions for large values of the arguments or parameters are often useful. In the preceding chapter we considered the asymptotic behavior of Sturm-Liouville and Bessel functions, restricting the variables to the real domain. In the present section we shall develop methods of obtaining representations which depend essentially on the use of complex variables and complex integration.

1. Stirling's Formula. As a first example of an asymptotic expansion we consider Stirling's formula. It will be derived by a method which will be used frequently in the future; here, however, complex integration does not occur. For $s > 0$ we have

$$\begin{aligned}\Gamma(s+1) &= \int_0^\infty t^s e^{-t}\,dt \\ &= s^{s+1}\int_0^\infty \tau^s e^{-s\tau}\,d\tau \qquad (t = s\tau) \\ &= s^{s+1}e^{-s}\int_0^\infty e^{-s(\tau-1-\log\tau)}\,d\tau \\ &= s^{s+1}e^{-s}\int_0^\infty e^{-sf(\tau)}\,d\tau \qquad (f(\tau) = \tau - 1 - \log\tau);\end{aligned}$$

the integrand has the value 1 for $\tau = 1$, for all other values it tends to zero with increasing s. Hence we may expect that only the immediate neighborhood of $\tau = 1$ will contribute essentially to the value of the integral when s is sufficiently large. Accordingly, we replace this integral by the integral over the interval $1 - \epsilon$ to $1 + \epsilon$ $(0 < \epsilon < \frac{1}{2})$, and begin by estimating the error incurred in neglecting the integrals from 0 to $1 - \epsilon$ and from $1 + \epsilon$ to ∞. For $\frac{1}{2} \leq \tau \leq 1$ we have

$$\begin{aligned}\tau - 1 - \log\tau &= \int_\tau^1 \left(\frac{1}{u} - 1\right) du \geq \int_\tau^1 (1-u)\,du \\ &= \tfrac{1}{2}(\tau-1)^2 \geq \tfrac{1}{8}(\tau-1)^2,\end{aligned}$$

and for $1 \leq \tau \leq 4$

$$\tau - 1 - \log\tau = \int_1^\tau \left(1 - \frac{1}{u}\right) du \geq \tfrac{1}{4}\int_1^\tau (u-1)\,du = \tfrac{1}{8}(\tau-1)^2.$$

In the integrals

$$\int_0^{1-\epsilon} e^{-sf(\tau)}\,d\tau, \qquad \int_{1+\epsilon}^{4} e^{-sf(\tau)}\,d\tau$$

we replace the integrand by its largest value, assumed at the points $1 \mp \epsilon$, and this in turn by the upper bound $e^{-s\epsilon^2/8}$. We thus obtain

$$\int_0^{1-\epsilon} + \int_{1+\epsilon}^{4} \leq 4e^{-s\epsilon^2/8}.$$

However, for $\tau \geq 4$ we have $\tau - 1 - \log \tau \geq \frac{3\tau}{4} - \log \tau > \frac{\tau}{4}$; hence for $s > 4$

$$\int_4^{\infty} e^{-s(\tau-1-\log\tau)}\,d\tau < \int_4^{\infty} e^{-s\tau/4}\,d\tau < e^{-s} < e^{-s\epsilon^2/8},$$

and setting $\epsilon = s^{-2/5}$,

$$e^s s^{-s-1}\Gamma(s+1) = \int_{1-\epsilon}^{1+\epsilon} e^{-sf(\tau)}\,d\tau + O(e^{-s^{1/5}/8}).^1$$

In order to find an approximation for the integral on the right, we make use of the relation

$$f(\tau) = \frac{(\tau-1)^2}{2} + (\tau-1)^3\psi(\tau),$$

where $\psi(\tau)$ is a function, regular in the interval $\frac{1}{2} \leq \tau \leq \frac{3}{2}$, whose absolute value in this interval does not exceed a finite bound M. From this relation, for $1 - \epsilon \leq \tau \leq 1 + \epsilon$, we conclude

$$e^{-s(\tau-1)^2/2}e^{-Ms^{-1/5}} \leq e^{-sf(\tau)} \leq e^{-s(\tau-1)^2/2}e^{Ms^{-1/5}}$$

and in addition

$$e^{-sf(\tau)} = e^{-s(\tau-1)^2/2}(1 + O(s^{-1/5})).$$

From this follows

$$\int_{1-\epsilon}^{1+\epsilon} e^{-sf(\tau)}\,d\tau = (1 + O(s^{-1/5})) \int_{-\epsilon}^{+\epsilon} e^{-su^2/2}\,du$$

$$= (1 + O(s^{-1/5})) \sqrt{\frac{2}{s}} \int_{-\epsilon\sqrt{s/2}}^{+\epsilon\sqrt{s/2}} e^{-v^2}\,dv$$

[1] The notation $O(g(s))$ has here the same meaning as in Ch. V, §11.

$$\int_{1-\epsilon}^{1+\epsilon} e^{-sf(\tau)}\,d\tau = \sqrt{\frac{2\pi}{s}}\,(1 + O(s^{-1/5}))(1 + O(e^{-s\epsilon^2/2}))$$
$$= \sqrt{\frac{2\pi}{s}}\,(1 + O(s^{-1/5})),$$

or in other words

$$\Gamma(s + 1) = s^{s+1/2}e^{-s}\sqrt{2\pi}(1 + O(s^{-1/5})); \tag{85}$$

hence we have

$$\Gamma(s + 1) \sim \sqrt{2\pi s}\,s^s e^{-s}. \tag{86}$$

2. Asymptotic Calculation of Hankel and Bessel Functions for Large Values of the Arguments. In a similar manner, we can obtain an asymptotic estimate for the Hankel functions $H_\lambda^1(z)$ for large $|z|$ inside an angle $-\pi/2 + \delta < \arg z < \pi/2 - \delta$ by considering the integral

$$H_\lambda^1(z) = \frac{\Gamma(\frac{1}{2} - \lambda)(z/2)^\lambda}{\pi i\Gamma(\frac{1}{2})} \int e^{iz\tau}(\tau^2 - 1)^{\lambda-1/2}\,d\tau$$

(see §2, 5); here the path of integration is given by the right path in Fig. 13, z must satisfy $-\pi/2 < \arg z < \pi/2$, and $\log(\tau^2 - 1)$ is assumed real for $\tau > 1$. Without changing the value of the integral, we can rotate the cuts in the τ-plane, together with the path of integration which encloses one of them, so that it has the direction $\pi/2 - \arg z$. If we then make the substitution

$$\tau - 1 = i\frac{u}{z}$$

the u-plane is slit by means of two cuts running horizontally from 0 to $2iz$, respectively, out to infinity, and the new path of integration surrounds the cut which runs along the positive real axis, going from right to left in the upper half-plane, in the lower half-plane from left to right. If by $u^{\lambda-1/2}$ we understand the uniquely determined branch in the cut plane which is positive on the lower edge of the positive real axis, and by $(1 + iu/2z)^{\lambda-1/2}$ the branch which assumes the value 1 for $u = 0$, we find

$$H_\lambda^1(z) = \frac{\Gamma(\frac{1}{2} - \lambda)}{\pi\sqrt{2\pi z}}\,e^{i(z+\pi\lambda/2-\pi/4)} \int e^{-u}u^{\lambda-1/2}\left(1 + \frac{iu}{2z}\right)^{\lambda-1/2} du.$$

Now if $\mathcal{R}_e\,(\lambda - \frac{1}{2}) > -1$, we can contract the loop around $u = 0$ and thus replace the integral over this loop by the integral taken over the lower edge of the positive real axis from 0 to ∞, minus the integral over the upper edge from ∞ to 0. But the latter equals the first integral times $e^{-2\pi i(\lambda+1/2)}$. Hence, using the supplementary formula for the gamma function, we have, after simple manipulation,

$$H_\lambda^1(z) = \left(\frac{2}{\pi z}\right)^{1/2} \frac{e^{i(z-\lambda\pi/2-\pi/4)}}{\Gamma(\lambda + \frac{1}{2})} \int_0^\infty e^{-u} u^{\lambda-1/2} \left(1 + \frac{ui}{2z}\right)^{\lambda-1/2} du. \tag{87}$$

Taylor's theorem with Cauchy's remainder term (denoted by R) gives us an expression for the last factor:

$$\begin{aligned}\left(1 + \frac{ui}{2z}\right)^{\lambda-1/2} &= \sum_{\nu=0}^{p-1} \binom{\lambda - \frac{1}{2}}{\nu} \left(\frac{ui}{2z}\right)^\nu \\ &\quad + p\binom{\lambda - \frac{1}{2}}{p} \left(\frac{ui}{2z}\right)^p \int_0^1 (1-t)^{p-1} \left(1 + \frac{tui}{2z}\right)^{\lambda-1/2-p} dt;\end{aligned} \tag{88}$$

observe that we thus obtain a useful estimate for the remainder.

Assuming that $\mathcal{R}_e\,(\lambda - \frac{1}{2} - p) < 0$, which is certainly true for sufficiently large p, we have for positive u

$$\left|1 + \frac{tui}{2z}\right| > \sin\delta, \qquad \left|\arg\left(1 + \frac{tui}{2z}\right)\right| < \pi,$$

$$\left|\left(1 + \frac{tui}{2z}\right)^{\lambda-1/2-p}\right| < e^{\pi|\mathcal{I}_{m(\lambda)}|} (\sin\delta)^{\mathcal{R}_e(\lambda-1/2-p)} = A_p,$$

where A_p is independent of z and t. Inserting (88) in (87) and integrating term by term, we obtain

$$\begin{aligned}H_\lambda^1(z) &= \left(\frac{2}{\pi z}\right)^{1/2} \frac{e^{i(z-\lambda\pi/2-\pi/4)}}{\Gamma(\lambda + \frac{1}{2})} \\ &\quad \cdot \left[\sum_{\nu=0}^{p-1} \binom{\lambda - \frac{1}{2}}{\nu} \Gamma(\lambda + \nu + \tfrac{1}{2}) \left(\frac{i}{2z}\right)^\nu + R_p\right],\end{aligned} \tag{89}$$

so that

$$|R_p| \leq A_p \left|p\binom{\lambda - \frac{1}{2}}{p} \left(\frac{i}{2z}\right)^p \left\{\int_0^1 (1-t)^{p-1}\, dt \int_0^\infty e^{-u} |u^{\lambda-1/2+p}|\, du\right\}\right|,$$

$$R_p = O(|z|^{-p}).$$

Again making the substitution $\tau + 1 = iu/z$, we obtain

$$H_\lambda^2(z) = \left(\frac{2}{\pi z}\right)^{1/2} \frac{e^{-i(z-\lambda\pi/2-\pi/4)}}{\Gamma(\lambda+\frac{1}{2})} \cdot \left[\sum_{\nu=0}^{p-1} \binom{\lambda-\frac{1}{2}}{\nu} \Gamma(\lambda+\nu+\tfrac{1}{2}) \left(-\frac{i}{2z}\right)^\nu + S_p\right], \quad S_p = O(|z|^{-p}), \tag{90}$$

and it follows that

$$\begin{aligned} J_\lambda(z) &= \tfrac{1}{2}(H_\lambda^1(z) + H_\lambda^2(z)) \\ &= \frac{1}{\Gamma(\lambda+\frac{1}{2})} \left(\frac{2}{\pi z}\right)^{1/2} \sum_{\nu=0}^{p-1} \binom{\lambda-\frac{1}{2}}{\nu} \frac{\Gamma(\lambda+\nu+\frac{1}{2})}{(2z)^\nu} \\ &\quad \cdot \begin{cases} (-1)^{\nu/2} \cos\left(z - \dfrac{\lambda\pi}{2} - \dfrac{\pi}{4}\right) \\ (-1)^{(\nu+1)/2} \sin\left(z - \dfrac{\lambda\pi}{2} - \dfrac{\pi}{4}\right) \end{cases} + O(|z|^{-p-1/2}), \end{aligned} \tag{91}$$

where the upper expression inside the braces refers to even ν, the lower to odd ν.

The first term of the expansion yields

$$J_\lambda(z) = \sqrt{\frac{2}{\pi z}} \cos\left(z - \frac{\lambda\pi}{2} - \frac{\pi}{4}\right) + O(|z|^{-3/2}), \tag{92}$$

which determines the limiting values of Ch. V, §11, 2:

$$\alpha_\infty = \sqrt{\frac{2}{\pi}}, \qquad \delta_\infty = -\frac{\lambda\pi}{2} - \frac{\pi}{4}.$$

3. The Saddle Point Method. In a great many cases a more general method, called the saddle point method, can be used to determine asymptotic values of integrals. Consider an integral

$$\int_C e^{zf(\tau)}\, d\tau$$

over a path of integration C on which the real part of $f(\tau)$ tends to $-\infty$ as we approach either end point. For large positive values of z the distant portions of the path of integration, in other words the parts corresponding to large negative real part $\mathcal{R}_e\, f(\tau)$, furnish a contribution which becomes smaller as z becomes larger. The path

of integration in the complex plane will now be deformed so that the part of it which, for large z, contributes significantly to the integral will be contracted to a neighborhood of a point. Thus we must choose a path of integration for which $\mathcal{R}_e f(\tau)$ falls off from a maximum value as fast as possible on each side. If we set $\tau = u + iv$ and think of the real part $\mathcal{R}_e f(\tau)$ as a surface lying over the u, v-plane—the surface has negative curvature at every point—our object will be attained provided it is possible to take the path over a saddle point or col in such a way that on both sides of the saddle point it falls off to large negative values of $\mathcal{R}_e f(\tau)$ as steeply as possible. In this case, for large positive values of z only the immediate neighborhood of the saddle point will be of significance.

The curves of steepest descent are given by the orthogonal trajectories of the level curves $\mathcal{R}_e f(\tau) = \text{const.}$, hence by the curves $\mathcal{I}_m f(\tau) = \text{const.}$ At a saddle point the derivatives of the functions $\mathcal{R}_e f(\tau)$ and $\mathcal{I}_m f(\tau)$ taken along the curve $\mathcal{I}_m f(\tau)$ vanish; therefore the derivative $f'(\tau)$ of the function $f(\tau)$ also vanishes and the saddle points will be found among the roots of the equation

$$f'(\tau) = 0.$$

The derivation of Stirling's formula is an example of this method; for, the real axis is the path that falls off most steeply from the saddle point $\tau = 1$.

4. Application of the Saddle Point Method to the Calculation of Hankel and Bessel Functions for Large Parameter and Large Argument. We shall use this method to evaluate the function (see formula (3), page 468)

$$H_\lambda^1(a\lambda) = -\frac{1}{\pi}\int_{L_1} e^{\lambda(-ia\sin\tau+i\tau)}\,d\tau$$

asymptotically for the case of real a and large positive λ. We separate the factor of λ in the exponent into its real and imaginary parts:

$$-ia\sin\tau + i\tau = f(\tau) = a\cos u\sinh v - v + i(u - a\sin u\cosh v).$$

The saddle points are roots of the equation $a\cos\tau = 1$ through which the curves $u - a\sin u\cosh v = \text{const.}$ should pass; we shall now try to combine them in such a way that suitable paths of integration are formed.

(*1*) If $a < 1$, say $a = 1/\cosh\alpha$ $(\alpha > 0)$, we have saddle points given

by $\tau = \pm \alpha i$ and the corresponding curves are $u - a \sin u \cosh v = 0$. They consist of the imaginary axis $u = 0$ and any branch through $\tau = \pm \alpha i$ which approaches the lines $u = \pm \pi$ from above and from below, respectively. This is shown in Figure 22, where the direction of increasing real part of $f(\tau)$ is indicated by arrows. The curve made up of the curves $\mathcal{I}_m(f) = 0$ traversed in an upward direction evidently yields H_λ^1; for, it may be deformed into L_1 except for a part beginning arbitrarily high which lies inside a strip $-\pi \leq u \leq -\pi + \epsilon$ and hence furnishes an arbitrarily small contribution to the value of the integral. The real part of $-ia \sin \tau + i\tau$ has its maximum $\alpha - \tanh \alpha$ for $\tau = -\alpha i$. Again we replace (see page 522) the path L_1 by the straight-line segment L' from $(-\alpha - \epsilon)i$ to $(-\alpha + \epsilon)i$ with $\epsilon = \lambda^{-2/5}$. If we then separate the remaining path of integration into two adjacent finite parts and two parts leading to infinity, we find an estimate which corresponds exactly to that derived in subsection 1, given by

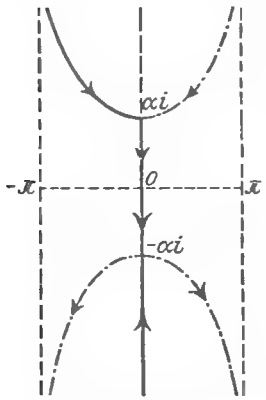


Figure 22

$$\int_{-i\infty}^{(-\alpha-\epsilon)i} e^{\lambda f(\tau)}\, d\tau + \int_{(-\alpha+\epsilon)i}^{-\pi+i\infty} e^{\lambda f(\tau)}\, d\tau = e^{\lambda(\alpha-\tanh\alpha)} O(e^{-c_1\lambda\epsilon^2}),$$

where c_1 (as well as c_2, c_3, etc. which will occur below) denotes a positive constant independent of λ (hence also of ϵ). That is, on the two finite intervals the absolute value of the integrand is at most equal to the values which it has at the points $(-\alpha \pm \epsilon)i$, and for these values the indicated approximation holds. On the infinite portions one easily finds a bound of the form $e^{-c\lambda(s+c')}$ for the absolute value of the integrand, where s is the arc length on the path of integration and c and c' are positive constants independent of ϵ and λ. For the contributions of these parts to the total integral this yields an estimate $O(e^{-c_1\lambda})$. On the portion L' itself, however,

$$|f(\tau) - (\alpha - \tanh\alpha + \tfrac{1}{2}f''(-\alpha i)(\tau + \alpha i)^2)| < c_2\epsilon^3,$$

$$f''(-\alpha i) = \tanh\alpha;$$

hence

$$e^{\lambda f(\tau)} = e^{\lambda[\alpha-\tanh\alpha+\tanh\alpha((\tau+\alpha i)^2/2)]}(1 + O(\lambda^{-1/5})),$$

$$\int_{(-\alpha-\epsilon)i}^{(-\alpha+\epsilon)i} e^{\lambda f(\tau)}\,d\tau = e^{\lambda(\alpha-\tanh\alpha)} \int_{(-\alpha-\epsilon)i}^{(-\alpha+\epsilon)i} e^{\lambda\tanh\alpha((\tau+\alpha i)^2/2)}\,d\tau(1 + O(\lambda^{-1/5}))$$

$$= i\sqrt{\frac{2}{\lambda\tanh\alpha}}\, e^{\lambda(\alpha-\tanh\alpha)} \int_{-\epsilon\sqrt{(\lambda\tanh\alpha)/2}}^{\epsilon\sqrt{(\lambda\tanh\alpha)/2}} e^{-u^2}\,du(1 + O(\lambda^{-1/5}))$$

$$= i\sqrt{\frac{2}{\lambda\tanh\alpha}}\, e^{\lambda(\alpha-\tanh\alpha)} \int_{-\infty}^{\infty} e^{-u^2}\,du(1 + O(e^{-c_3\lambda\epsilon^2}))(1 + O(\lambda^{-1/5}))$$

$$= i\sqrt{\frac{2\pi}{\lambda\tanh\alpha}}\, e^{\lambda(\alpha-\tanh\alpha)}(1 + O(e^{-c_3\lambda\epsilon^2}))(1 + O(\lambda^{-1/5})).$$

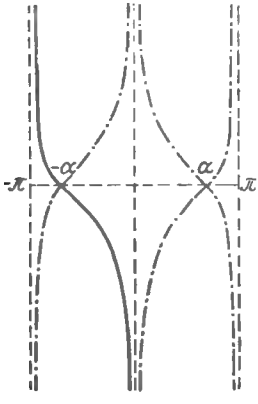


Figure 23

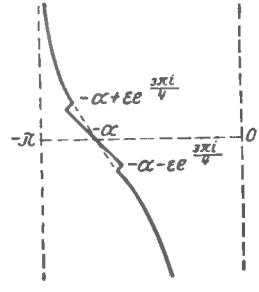


Figure 24

Thus we have

$$H_\lambda'(a\lambda) = -i\sqrt{\frac{2}{\pi\lambda\tanh\alpha}}\, e^{\lambda(\alpha-\tanh\alpha)}(1 + O(\lambda^{-1/5})). \tag{93}$$

(2) If $a > 1$, say $a = 1/\cos\alpha$ $(0 < \alpha < \pi/2)$, we have the saddle points $\tau = \pm\alpha$ and the curves

$$u - a\sin u\cosh v = \pm(\alpha - a\sin\alpha), \qquad \cosh v = \frac{u \mp (\alpha - \tan\alpha)}{a\sin u}$$

which are reproduced in Figure 23; the solid path represents $H_\lambda^1(x)$. In the neighborhood of the saddle point, we replace this path by a line-segment making an angle $-\pi/4$ with the real axis, and by connecting segments of bounded length on which $\mathcal{I}_m f(\tau)$ does not assume values larger than its value at the points $\tau = -\alpha \pm \epsilon e^{3\pi i/4}$ (see Figure 24). Again we set $\epsilon = \lambda^{-2/5}$, and obtain as before

$$\int_{L_1} e^{\lambda f(\tau)}\,d\tau = e^{\lambda f(-\alpha)} \int_{-\alpha-\epsilon e^{3\pi i/4}}^{-\alpha+\epsilon e^{3\pi i/4}} e^{(\lambda/2) f''(-\alpha)(\tau+\alpha)^2}\,d\tau(1 + O(\lambda^{-1/5}))$$

$$= e^{i\lambda(\tan\alpha-\alpha)} \int_{-\alpha-\epsilon e^{3\pi i/4}}^{-\alpha+\epsilon e^{3\pi i/4}} e^{(-\lambda i/2)\tan\alpha(\tau+\alpha)^2}\,d\tau(1 + O(\lambda^{-1/5}))$$

$$= e^{3\pi i/4}\sqrt{\frac{2}{\lambda\tan\alpha}}\, e^{i\lambda(\tan\alpha-\alpha)} \int_{-\epsilon\sqrt{(\lambda\tan\alpha)/2}}^{+\epsilon\sqrt{(\lambda\tan\alpha)/2}} e^{-u^2}\,du(1 + O(\lambda^{-1/5}))$$

$$= e^{3\pi i/4}\sqrt{\frac{2\pi}{\lambda\tan\alpha}}\, e^{i\lambda(\tan\alpha-\alpha)}\,(1 + O(e^{-c_3\lambda\epsilon^2}))(1 + O(\lambda^{-1/5})),$$

$$H_\lambda^1(a\lambda) = -e^{3\pi i/4}\sqrt{\frac{2}{\pi\lambda\tan\alpha}}\, e^{i\lambda(\tan\alpha-\alpha)}(1 + O(\lambda^{-1/5})). \tag{94}$$

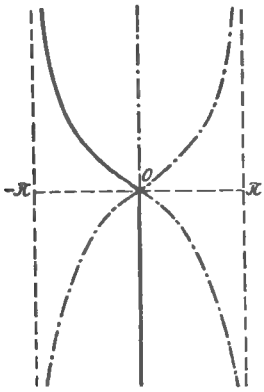


Figure 25

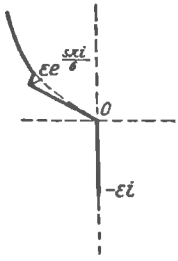


Figure 26

(3) If $a = 1, f''(\tau)$ also vanishes at the saddle point $\tau = 0$ and the curve $\mathcal{I}_m\, f(\tau) = u - \sin u \cosh v = \mathcal{I}_m\, f(0) = 0$ has three branches through $\tau = 0$ (Figure 25), one of which is the imaginary axis. Again we replace the curved part of the path L_1 (Figure 25) which is close to $\tau = 0$ by a straight segment of length $\epsilon = \lambda^{-1/4}$ inclined at the angle $5\pi/6$ with the real axis (see Figure 26) and obtain for all τ of the path of integration between $-\epsilon i$ and $\epsilon e^{5\pi i/6}$

$$\left| f(\tau) - \frac{i\tau^3}{6} \right| \le c_1\epsilon^5.$$

Moreover, we have

$$\int_{L_1} e^{\lambda f(\tau)}\,d\tau = \int_{-\epsilon i}^{\epsilon e^{5\pi i/6}} e^{\lambda f(\tau)}\,d\tau + O(e^{-c_1\lambda\epsilon^3}),$$

$$\int_{-\epsilon i}^{\epsilon e^{5\pi i/6}} e^{\lambda f(\tau)}\, d\tau = \int_{-\epsilon i}^{\epsilon e^{5\pi i/6}} e^{\lambda i\tau^3/6}\, d\tau(1 + O(\lambda^{-1/4})),$$

$$\int_{-\epsilon i}^{\epsilon e^{5\pi i/6}} e^{\lambda i\tau^3/6}\, d\tau = \int_0^{\epsilon e^{5\pi i/6}} - \int_0^{-\epsilon i} = \sqrt[3]{\frac{6}{\lambda}}\,(e^{5\pi i/6} + i) \int_0^{\epsilon\sqrt[3]{\lambda/6}} e^{-u^3}\, du;$$

in the last transformation we use the substitution $\tau = \sqrt[3]{6/\lambda}\, e^{5\pi i/6} u$ in the first integral and $\tau = -\sqrt[3]{6/\lambda}\, iu$ in the second. The right side of the last equation equals

$$\sqrt[3]{\frac{6}{\lambda}}\,(e^{5\pi i/6} + i) \int_0^{\infty} e^{-u^3}\, du(1 + O(e^{-c_3\epsilon^3\lambda})),$$

if $\epsilon^3\lambda$ remains greater than a positive bound. On the other hand we have

$$\int_0^{\infty} e^{-u^3}\, du = \tfrac{1}{3}\int_0^{\infty} e^{-t} t^{-2/3}\, dt = \tfrac{1}{3}\Gamma(\tfrac{1}{3})$$

and, therefore, finally

$$H_\lambda^1(\lambda) = -\frac{1}{3\pi}\,\Gamma(\tfrac{1}{3})(e^{5\pi i/6} + i)\sqrt[3]{\frac{6}{\lambda}}\,(1 + O(\lambda^{-1/4})). \tag{95}$$

To find asymptotic formulas for $J_\lambda(a\lambda)$ in the cases $a \geq 1$ we use the formulas derived above for $H_\lambda^1(a\lambda)$ and the following three formulas which are found in a similar way:

$$H_\lambda^2(a\lambda) = i\sqrt{\frac{2}{\pi\lambda \tanh \alpha}}\, e^{\lambda(\alpha - \tanh \alpha)}(1 + O(\lambda^{-1/5})) \quad (a < 1), \tag{96}$$

$$H_\lambda^2(a\lambda) = -e^{-3\pi i/4}\sqrt{\frac{2}{\pi\lambda \tan \alpha}}\, e^{-i\lambda(\tan \alpha - \alpha)}(1 + O(\lambda^{-1/5})) \quad (a > 1), \tag{96'}$$

$$H_\lambda^2(\lambda) = -\frac{1}{3\pi}\,\Gamma(\tfrac{1}{3})(e^{-5\pi i/6} - i)\sqrt[3]{\frac{6}{\lambda}}\,(1 + O(\lambda^{-1/4})) \quad (a = 1). \tag{96''}$$

We combine them by means of

$$J_\lambda(x) = \tfrac{1}{2}(H_\lambda^1(x) + H_\lambda^2(x));$$

the principal term turns out to be zero only in the case $a < 1$. In this case we can take the path shown in Figure 27 for J_λ also, ob-

taining by the same method

$$J_\lambda(a\lambda) = \frac{2}{\sqrt{2\pi\lambda \tanh \alpha}}\, e^{\lambda(\tanh \alpha - \alpha)}(1 + O(\lambda^{-1/5})).$$

5. General Remarks on the Saddle Point Method. Here the saddle point method has been used only to derive asymptotic formulas for the first terms of asymptotic series found as originally indicated. The reader is referred to the detailed presentation of these series in G. N. Watson, A Treatise on the Theory of Bessel Functions, Cambridge University Press, Cambridge, 1922, and to the literature, in particular P. Debye, Näherungsformeln für die Zylinderfunktionen für grosse Werte des Arguments und unbeschränkt veränderliche Werte des Index, Math. Ann., Vol. 67, 1909, pp. 535–558.

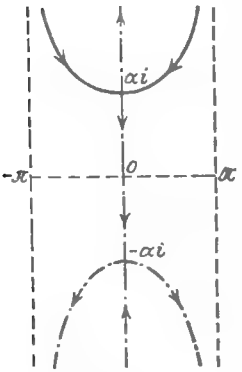


Figure 27

In applications of the saddle point method it is not necessary to take the path of integration precisely in the manner described. It is enough if eventually, i.e. for large values of the parameter in terms of which the functions are expanded, the path comes sufficiently close to the one described. In this way Faber[1] obtains a number of asymptotic series, e.g. for Hermite and Laguerre polynomials.

6. The Darboux Method. A different method for deriving asymptotic formulas is due to Darboux.[2] Let the quantities a_ν in question be given as the coefficients of a power series, hence by a generating function $K(\zeta) = \sum_{\nu=0}^{\infty} a_\nu \zeta^\nu$. If we know the singularities of this function on the circle of convergence—say $|\zeta| = 1$, $\zeta = e^{i\varphi}$—and if by subtracting known functions $f_n(\zeta) = \sum_{\nu=0}^{\infty} \alpha_{n\nu}\zeta^\nu$ we can insure that the remainder $K - f_n$ converges uniformly, as we approach the circle of convergence, to an n-times continuously differentiable func-

[1] G. Faber, Abschätzung von Funktionen grosser Zahlen, S.-Ber. Akad. München (math.-phys. Kl.), 1922, pp. 285–304.

[2] G. Darboux, Mémoire sur l'approximation des fonctions de très-grands nombres, et sur une classe étendue de développements en série, J. de math. pures et appl., Ser. 3, Vol. 4, 1878, pp. 5–56 and 377–416. See also A. Haar, Über asymptotische Entwicklungen von Funktionen, Math. Ann., Vol. 96, 1926, pp. 69–107.

tion of φ, then the coefficients $a_\nu - \alpha_{n\nu}$ of the power series

$$K(\zeta) - f_n(\zeta) = \sum_{\nu=0}^{\infty} (a_\nu - \alpha_{n\nu})\zeta^\nu$$

are the Fourier coefficients of an n-times continuously differentiable (i.e. for $n = 0$, just continuous) function of φ; hence, by Ch. II, §5, 3, they satisfy the condition

$$\lim_{\nu\to\infty} \nu^{n-1} \mid a_\nu - \alpha_{n\nu} \mid = 0.$$

Thus if ν is large, the approximation of a_ν by $\alpha_{n\nu}$ becomes better as n increases.

7. Application of the Darboux Method to the Asymptotic Expansion of Legendre Polynomials. We apply the method to the Legendre polynomials $P_\nu(x)$ which are given by means of the generating function

$$K(z,\zeta) = \frac{1}{\sqrt{1 - 2z\zeta + \zeta^2}} = \sum_{\nu=0}^{\infty} P_\nu(z)\zeta^\nu. \tag{97}$$

Assume $-1 < z < 1$, $z = \cos\varphi$, $0 < \varphi < \pi$. Then $1 - 2z\zeta + \zeta^2 = (\zeta - e^{\varphi i})(\zeta - e^{-\varphi i})$; the circle of convergence has radius 1, and on it lie the singular points $\zeta = e^{\pm\varphi i}$. In order to derive the series expansion of K according to powers of $\zeta - e^{\pm\varphi i}$, we make the convention that

$$\sqrt{\zeta - e^{\pm\varphi i}} = e^{\pm i(\varphi+\pi)/2}\sqrt{1 - \zeta e^{\mp\varphi i}},$$

where the square root on the right denotes the branch presented by the binomial series.[1] We obtain

$$\begin{aligned} K(z,\zeta) &= \frac{1}{\sqrt{\zeta - e^{\varphi i}}}\left[(\zeta - e^{\varphi i}) + (e^{\varphi i} - e^{-\varphi i})\right]^{-1/2} \\ &= \frac{e^{3\pi i/4}}{\sqrt{2\sin\varphi}} \frac{1}{\sqrt{\zeta - e^{\varphi i}}} \sum_{\nu=0}^{\infty} \binom{-\frac{1}{2}}{\nu} \left(\frac{\zeta - e^{\varphi i}}{e^{\varphi i} - e^{-\varphi i}}\right)^\nu \\ &= \frac{e^{-3\pi i/4}}{\sqrt{2\sin\varphi}} \frac{1}{\sqrt{\zeta - e^{-\varphi i}}} \sum_{\nu=0}^{\infty} \binom{-\frac{1}{2}}{\nu} \left(\frac{\zeta - e^{-\varphi i}}{e^{-\varphi i} - e^{\varphi i}}\right)^\nu. \end{aligned}$$

[1] Thus if a is a positive number, for $\zeta = e^{\varphi i} - a$ the root $\sqrt{\zeta - e^{\varphi i}}$ is positive imaginary, while for $\zeta = e^{-\varphi i} - a$ the root $\sqrt{\zeta - e^{-\varphi i}}$ is negative imaginary. The above convention is in agreement with the requirement, which follows from formula (97), that for $\zeta = 0$ the root $\sqrt{1 - 2z\zeta + \zeta^2}$ should take the value $+1$.

If we set

$$f_n(z,\zeta) = \frac{1}{\sqrt{2\sin\varphi}} \sum_{\nu=0}^{n} \binom{-\frac{1}{2}}{\nu} \cdot \left\{ e^{3\pi i/4} \frac{(\zeta - e^{\varphi i})^{\nu-1/2}}{(e^{\varphi i} - e^{-\varphi i})^{\nu}} + e^{-3\pi i/4} \frac{(\zeta - e^{-\varphi i})^{\nu-1/2}}{(e^{-\varphi i} - e^{\varphi i})^{\nu}} \right\},$$

then $K - f_n$ is n-times continuously differentiable on the circle of convergence. Hence if we expand f_n in powers of ζ and, for convenience, write $(3\pi/4) + (\varphi + \pi)(\nu - \frac{1}{2}) = \omega$, we obtain

$$\begin{aligned} f_n(z,\zeta) &= \frac{1}{\sqrt{2\sin\varphi}} \sum_{\nu=0}^{n} \binom{-\frac{1}{2}}{\nu} \left\{ \frac{e^{i\omega}(1 - \zeta e^{-\varphi i})^{\nu-1/2}}{(e^{\varphi i} - e^{-\varphi i})^{\nu}} + \frac{e^{-i\omega}(1 - \zeta e^{\varphi i})^{\nu-1/2}}{(e^{-\varphi i} - e^{\varphi i})^{\nu}} \right\} \\ &= \frac{1}{\sqrt{2\sin\varphi}} \sum_{\nu=0}^{n} \binom{-\frac{1}{2}}{\nu} \sum_{\mu=0}^{\infty} \binom{\nu - \frac{1}{2}}{\mu} \zeta^{\mu} \cdot \\ &\qquad \left\{ \frac{e^{i[\omega - (\varphi+\pi)\mu]}}{(e^{\varphi i} - e^{-\varphi i})^{\nu}} + \frac{e^{-i[\omega - (\varphi+\pi)\mu]}}{(e^{-\varphi i} - e^{\varphi i})^{\nu}} \right\} \\ &= \sum_{\mu=0}^{\infty} p_{n\mu}(z)\zeta^{\mu} \end{aligned}$$

with

$$p_{n\mu} = \frac{1}{\sqrt{2\sin\varphi}} \sum_{\nu=0}^{n} \binom{-\frac{1}{2}}{\nu} \binom{\nu - \frac{1}{2}}{\mu} \frac{1}{(2\sin\varphi)^{\nu}} \cdot$$

$$[\exp\{i[\pi/4 + (\nu - \mu - \tfrac{1}{2})\varphi - (\mu - \nu/2)\pi]\}$$

$$+ \exp\{-i[\pi/4 + (\nu - \mu - \tfrac{1}{2})\varphi - (\mu - \nu/2)\pi]\}],$$

$$(98) \qquad p_{n\mu} = \frac{2}{\sqrt{2\sin\varphi}} \sum_{\nu=0}^{n} \binom{-\frac{1}{2}}{\nu} \binom{\nu - \frac{1}{2}}{\mu} \frac{(-1)^{\mu}}{(2\sin\varphi)^{\nu}} \cos\left(\frac{\pi}{4}(1 + 2\nu) - (\nu - \mu - \tfrac{1}{2})\varphi\right),$$

we obtain

$$P_{\mu}(z) = p_{n,\mu}(z) + O(\mu^{-n})$$

uniformly in every interval of the form $-1 + \epsilon \leq z \leq 1 - \epsilon$ $(0 < \epsilon < 1)$. In view of the fact that $p_{n+1,\mu} - p_{n,\mu} = O(\mu^{-n-1})$, it follows that

$$P_{\mu}(z) = p_{n,\mu}(z) + O(\mu^{-n-1}).$$

The first term of this asymptotic expansion may be written as

$$P_\mu(z) = \frac{2}{\sqrt{2 \sin \varphi}} \frac{1 \cdot 3 \cdots (2\mu - 1)}{2 \cdot 4 \cdots 2\mu} \cos\left(\frac{\pi}{4} - (\mu + \tfrac{1}{2})\varphi\right) + O\left(\frac{1}{\mu}\right). \tag{99}$$

If z is not real between -1 and $+1$, then since the singular points ζ_1, ζ_2 satisfy $\zeta_1\zeta_2 = 1$, one of them, say ζ_1, has a modulus $|\zeta_1| < 1$, the other a modulus $|\zeta_2| > 1$. Only the singular point ζ_1 lies on the circle of convergence $|\zeta| = |\zeta_1|$; hence we need consider only this singularity. Accordingly, if we transform the first n terms of the expansion of $K(z, \zeta)$ in terms of powers of $\zeta - \zeta_1$ into a power series in ζ, the coefficients provide asymptotic expressions for $P_\nu(z)$; however, now only

$$|\zeta_1|^n(P_\mu - p_{n\mu}) = O(\mu^{-n-1})$$

holds.

Appendix to Chapter VII

§7. *Transformation of Spherical Harmonics*[1]

1. Introduction and Notation

Let x, y, z be Cartesian coordinates and let r, θ, φ be spherical polar coordinates defined by

$$x = r \cos\varphi \sin\theta, \qquad y = r \sin\varphi \sin\theta, \qquad z = r \cos\theta. \tag{100}$$

Let Δ denote Laplace's operator

$$\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} = \frac{\partial^2}{\partial r^2} + \frac{2}{r}\frac{\partial}{\partial r} + \frac{1}{r^2}\frac{\partial^2}{\partial \theta^2} + \frac{\cot\theta}{r^2}\frac{\partial}{\partial \theta} + \frac{1}{r^2 \sin^2\theta}\frac{\partial^2}{\partial \varphi^2}. \tag{101}$$

A function u which satisfies $\Delta u = 0$ is called harmonic. If $p_n(\theta, \varphi)$ is a one-valued function of x, y, z such that $r^n p_n$ is a harmonic func-

[1] The author is indebted to W. Magnus for this appendix, which is based on unpublished notes by the late G. Herglotz.

tion, p_n is called a spherical harmonic of order n. It was observed in §5 that $r^n p_n$ is a homogeneous harmonic polynomial of degree n in x, y, z and vice versa.

In several problems of mathematical physics the question arises how any set of linearly independent spherical harmonics is transformed if the coordinate system is rotated but the origin $r = 0$ kept fixed. To answer this question, we shall characterize the orthogonal transformations in terms of independent parameters q_ν ($\nu = 1, 2, 3, 4$). Coordinate rotation in a system of $2n + 1$ linearly independent spherical harmonics of order n is expressed by a linear substitution whose coefficients form a complete system of orthogonal hyperspherical harmonics of order $2n$ in the four-dimensional space of the parameters q_ν.

There exists no analogue to this result in the case of more than three variables.

2. Orthogonal Transformations

Let O be an orthogonal matrix of three rows and three columns; O is characterized by

$$OO' = I \tag{102}$$

where O' is the transposed matrix of O and I denotes the identity or unit matrix. We shall consider only "proper" orthogonal matrices; i.e., we assume that the determinant $|O|$ of O equals $+1$. Let

$$A = \begin{pmatrix} 0 & q_3 & -q_2 \\ -q_3 & 0 & q_1 \\ q_2 & -q_1 & 0 \end{pmatrix} \tag{103}$$

be a skew-symmetric matrix, i.e. one for which $A' = -A$. Let q_4 be a fourth parameter and let

$$w = (q_1^2 + q_2^2 + q_3^2)^{1/2}, \qquad v = (q_1^2 + q_2^2 + q_3^2 + q_4^2)^{1/2}. \tag{104}$$

Then we have

CAYLEY'S THEOREM: *For any real values of the q_ν ($\nu = 1, 2, 3, 4$) for which $v > 0$, the matrix*

$$O = (q_4 I + A)(q_4 I - A)^{-1} = \frac{1}{v^2}(v^2 I + 2q_4 A + 2A^2), \tag{105}$$

or

$$(106)\quad O = \frac{1}{v^2}\begin{pmatrix} q_4^2+q_1^2-q_2^2-q_3^2, & 2q_1q_2 - 2q_3q_4, & 2q_1q_3 + 2q_2q_4 \\ 2q_1q_2 + 2q_3q_4, & q_4^2+q_2^2-q_1^2-q_3^2, & 2q_2q_3 - 2q_1q_4 \\ 2q_1q_3 - 2q_2q_4, & 2q_2q_3 + 2q_1q_4, & q_4^2+q_3^2-q_1^2-q_2^2 \end{pmatrix},$$

is orthogonal; q_1/w, q_2/w, q_3/w *are the direction cosines of the axis of rotation with respect to the x-, y-, and z-axes. If* ω *denotes the angle of rotation, then*

$$(107)\qquad \cos\frac{\omega}{2} = \frac{q_4}{v}, \qquad \sin\frac{\omega}{2} = \frac{w}{v}.$$

We normalize the parameters q_ν *in such a way that*

$$(108)\qquad v^2 = q_1^2 + q_2^2 + q_3^2 + q_4^2 = 1$$

holds; then any two sets q_ν *and* q_ν^* *of values of the parameters define the same O if and only if* $q_\nu = q_\nu^*$ *or* $q_\nu = -q_\nu^*$ *for* $\nu = 1, 2, 3, 4$. If $w = 0$, every direction in the space remains fixed and O becomes the identity I.

Proof: We say that two matrices B and C commute if $BC = CB$. We may then perform all algebraic operations as if B and C were numbers provided we do not divide by any matrix unless its determinant is different from zero; for instance, if $|B| \neq 0$, B^{-1} and C also commute. Using this fact and the well-known relation $(B')^{-1} = (B^{-1})'$ (where the prime denotes a transposed matrix) we see at once: If $q_4 I - A$ has an inverse and if, as in (105), we set

$$(109)\qquad O = (q_4 I + A)(q_4 I - A)^{-1},$$

then

$$(110)\quad O' = (q_4 I' - A')^{-1}(q_4 I' + A') = (q_4 I - A)(q_4 I + A)^{-1} = O^{-1}.$$

Therefore O is orthogonal. Similarly, we find for any orthogonal O for which $|I + O| \neq 0$ that

$$(111)\qquad A = q_4(I - O)(I + O)^{-1}$$

is a skew-symmetric matrix. But (111) is exactly the expression for A which we obtain if we "solve" (109) with respect to A. This

proves the first part of Cayley's theorem (i.e. (105) and (106)) unless $|I + O| = 0$. Even in this case, the right side in (109) has a meaning if we write it in a different form. Since every matrix satisfies its characteristic equation, we have

$$A^3 + w^2 A = 0, \tag{112}$$

which (for $v > 0$) leads to

$$q_4 I + A = v^{-2}(v^2 I + 2q_4 A + 2A^2)(q_4 I - A). \tag{113}$$

We may define the right side in (109) as $v^{-2}(v^2 I + 2q_4 A + 2A^2)$; then, however, we have to show that this is an orthogonal matrix even if $|q_4 I - A| = 0$, i.e. if q_4 is an eigenvalue of A. Since A is skew-symmetric, its eigenvalues are pure imaginary. Hence $q_4 = 0$ and

$$v^{-2}(v^2 I + 2q_4 A + 2A^2) = I + 2w^{-2}A^2; \tag{114}$$

because of (112) we have

$$(I + 2w^{-2}A^2)^2 = I + 4w^{-2}A^2 + 4w^{-4}A^4 = I. \tag{115}$$

Therefore, for $q_4 = 0$, $I + 2w^{-2}A^2$ is a symmetric orthogonal matrix. It can be verified that all orthogonal matrices of this type can be written in the form $I + 2w^{-2}A^2$. On the other hand, any orthogonal matrix for which $|O| = +1$ and $|I + O| = 0$ is necessarily symmetric because such an O must have the eigenvalues $+1$, -1, -1; its square has the eigenvalues 1, 1, 1, and is therefore the identity. But $O^2 = OO' = I$ implies that O is symmetric.

In order to interpret the q_ν geometrically, we observe first that the vector (q_1, q_2, q_3) is an eigenvector of A belonging to the eigenvalue zero. It follows from (105) that (q_1, q_2, q_3) is an eigenvector of O belonging to the eigenvalue 1 and therefore defining the axis of rotation. Since the eigenvalues of O are $\exp\{i\omega\}$, $\exp\{-i\omega\}$, and 1. and since the sum of these is the trace of O, (107) follows from (106) except for the sign of $\cos(\omega/2)$, which need not be determined; for if we change $\omega/2$ by π, then ω changes by 2π.

Apparently, Cayley's theorem can be extended to more than three dimensions, but then the explicit formulas (105) and (106) become more complicated. In three dimensions, (106) can be stated in a

simple manner by using quaternions.[1] Let the vectors $\mathbf{x}$ and $\mathbf{x}'$ be defined by $\mathbf{x} = (x, y, z)$ and $\mathbf{x}' = O\mathbf{x}$. Let ξ, ξ', α be quaternions defined by

$$\xi = xi + yj + zk, \qquad \xi' = x'i + y'j + z'k \tag{116}$$

and

$$\alpha = -q_4 + q_1 i + q_2 j + q_3 k; \tag{116'}$$

then

$$\xi' = \alpha \xi \alpha^{-1} \tag{117}$$

where $\alpha\alpha^{-1} = \alpha^{-1}\alpha = 1$ and $\alpha^{-1} = v^{-2}(-q_4 - q_1 i - q_2 j - q_3 k)$.

It may be observed that there exists a one-to-two correspondence between the elements of the orthogonal group and the points on a four-dimensional hypersphere. There also exists a group G such that the orthogonal group is a quotient group of G and the elements of G are in a one-to-one correspondence with the points of the hypersphere. For a definition of this group see §4; for its significance consult, e.g., B. L. van der Waerden, *Die gruppentheoretische Methode in der Quantenmechanik*, §16, J. Springer, Berlin, 1932.

3. A Generating Function for Spherical Harmonics

Let $P_{n,l}(x)$ denote the functions defined in Ch. VII, §5, 1 for $l = 0, 1, 2, \cdots, n$ and let

$$P_{n,-l}(x) = (-1)^l \frac{(n-l)!}{(n+l)!} P_{n,l}(x). \tag{118}$$

Then we know that the $2n + 1$ functions

$$P_{n,l}(\cos\theta) e^{il\varphi} \qquad (l = 0, \pm 1, \pm 2, \cdots, \pm n) \tag{119}$$

[1] Quaternions are "hypercomplex" numbers with three "imaginary" units i, j, k which satisfy the relations

$$i^2 = j^2 = k^2 = -1,$$

$$ij = k, \quad jk = i, \quad ki = j,$$

$$ij + ji = 0, \quad ik + ki = 0, \quad jk + kj = 0.$$

The general quaternion is $\alpha = a + bi + cj + dk$ where a, b, c, d are real. All laws of algebra with the exception of the commutative law of multiplication are satisfied by quaternions.

form a complete set of linearly independent spherical harmonics of order n. We assert that:

The homogeneous polynomials $h_{n,l}$ of degree n in x, y, z which are defined by

$$\{x - iy - 2zt - (x + iy)t^2\}^n = t^n \sum_{l=-n}^{+n} h_{n,l}(x, y, z)t^l \tag{120}$$

form a complete system of linearly independent harmonic polynomials of degree n. If spherical polar coordinates are introduced by equation (100), *we have*

$$h_{n,l}(x, y, z) = r^n p_{n,l}(\theta, \varphi) \tag{121}$$

where

$$p_{n,l}(\theta, \varphi) = (-1)^n \frac{2^n\, n!}{(n + l)!} P_{n,l}(\cos\theta) e^{il\varphi}. \tag{122}$$

If x, y, z are connected by a relation

$$x^2 + y^2 + z^2 = 0, \tag{123}$$

then $h_{n,l}$ can be expressed in terms of two parameters w_1, w_2 by

$$h_{n,l}(x, y, z) = (-1)^n \binom{2n}{n + l} w_1^{n+l} w_2^{n-l} \tag{124}$$

where

$$x + iy = w_1^2, \qquad -x + iy = w_2^2, \qquad z = w_1 w_2\,. \tag{125}$$

Proof: If a, b, c are any three numbers which satisfy $a^2 + b^2 + c^2 = 0$, then clearly $\Delta(ax + by + cz)^n = 0$ holds; substituting $a = 1 - t^2$, $b = -i(1 + t^2)$, $c = -2t$, we obtain (120). In order to prove equations (121) and (122) we substitute t^{-1} for t in (120) and find

$$h_{n,-l} = (-1)^l \bar{h}_{n,l} \tag{126}$$

where the bar denotes a conjugate complex quantity. Therefore it suffices to prove (122) for $l \geq 0$. Introducing $t = e^{-i\varphi}s$ in (120), we find

$$\{(1 - s^2)\sin\theta - 2s\cos\theta\}^n = s^n \sum_{l=-n}^{+n} p_{n,l}(\theta, \varphi) e^{-il\varphi} s^l \tag{127}$$

and see that

$$p_{n,l}e^{-il\varphi} = f_{n,l}(\theta)$$

is a real function of θ only. From (126) we obtain

$$f_{n,-l}(\theta) = (-1)^l f_{n,l}(\theta). \tag{128}$$

If we multiply (127) by s^{-n}, substitute $\exp\{i(\omega + \pi/2)\}$ for s, and use (128), we obtain

$$(\cos\theta + i\cos\omega\sin\theta)^n = (-2)^{-n}\left\{f_{n,0} + 2\sum_{l=1}^{n} i^l f_{n,l}\cos l\omega\right\}. \tag{129}$$

On the other hand, the integral representations for $P_{n,l}(x)$ with $l = 0, 1, 2, \cdots, n$ (see §4, 4) show that

$$a_l = \frac{i^l}{(n+l)!}P_{n,l}(x)$$

is the Fourier coefficient in the expansion of

$$(x + i\sqrt{1-x^2}\cos\omega)^n = a_0 + 2\sum_{l=1}^{n} a_l\cos l\omega. \tag{130}$$

Setting $x = \cos\theta$ and comparing equations (130) and (129), we see that (122) holds.

Finally, (124) becomes a trivial application of the binomial theorem if in (120) we substitute the expressions for x, y, z given by (125).

It should be emphasized that equation (123) is not only a consequence of (125), but that any three numbers x, y, z satisfying (123) define exactly two pairs of numbers w_1, w_2 such that equations (125) hold. Since $x + iy$ and $x - iy$ are algebraically independent, the polynomials in w_1, w_2 on the right side of (124) are linearly independent. Therefore, (125) maps the harmonic polynomials $h_{n,l}$ on the linearly independent polynomials (of even degree) in w_1 and w_2. This also shows that every harmonic polynomial can be characterized by exactly one polynomial in three variables the sum of the squares of which is zero. Earlier in this chapter, the same fact was established by showing that the harmonic polynomials can be obtained if we apply a polynomial in the differential operators $\partial/\partial x$, $\partial/\partial y$, $\partial/\partial z$ to $r^{-1} = (x^2 + y^2 + z^2)^{-1/2}$, where the sum of the squares of these operators (if applied to r^{-1}) is evidently zero.

4. Transformation Formula

The significance of the formulas (120) to (125) will become clear when we have shown that any orthogonal transformation of x, y, z can be obtained from a linear substitution of w_1 and w_2. This expresses the well-known fact that the real orthogonal group in three variables has a representation by (complex) two-by-two matrices.

Lemma: Let w_1, w_2, x, y, z be connected by (125) and assume that

$$w_1' = \alpha w_1 + \beta w_2, \qquad w_2' = -\bar{\beta} w_1 + \bar{\alpha} w_2 \tag{131}$$

and

$$x' + iy' = w_1'^2, \qquad -x' + iy' = w_2'^2, \qquad z' = w_1' w_2'. \tag{132}$$

Then (125), (131), (132) define a linear substitution M by which the vector $\mathbf{x} = (x, y, z)$ is mapped on $\mathbf{x}' = (x', y', z') = M\mathbf{x}$. If we choose

$$\alpha = (q_4 - iq_3)/v, \qquad \beta = (-q_2 + iq_1)/v, \tag{133}$$

then M is the matrix O defined by (106).

The proof is trivial. If we replace α, β by $-\alpha, -\beta$ we obtain the same matrix $M = O$ as before. This shows that there are (at least) two different substitutions (131) which lead to the same orthogonal substitution O. As a matter of fact, there are exactly two such substitutions (132), and the group of these is the group G mentioned at the end of subsection 2 of this Appendix.

We may use this lemma to prove the following theorem which in its present form is due to G. Herglotz:

Let $\mathbf{x} = (x, y, z)$ be transformed into $\mathbf{x}' = (x', y', z')$ by the orthogonal transformation O defined by (105), (106). *Then the spherical harmonics $h_{n,l}(x', y', z')$ can be expressed in terms of the $h_{n,l}(x, y, z)$ (which are defined by* (120)) *by*

$$h_{n,l}(x', y', z') = \sum_{r=-n}^{n} \frac{\binom{2n}{l+n}}{\binom{2n}{r+n}} v^{-2n} H_{2n}^{(n+l,n+r)}(q_\mu) h_{n,r}(x, y, z) \tag{134}$$

where the $H_{2n}^{(n+l,n+r)}$ are homogeneous polynomials of degree $2n$ in the variables q_μ $(\mu = 1, 2, 3, 4)$ which satisfy

$$\sum_{\mu=1}^{4} \frac{\partial^2}{\partial q_\mu^2} H_{2n}^{(n+l,n+r)}(q_\mu) = 0 \tag{135}$$

and can be defined in terms of the generating function

$$G_{2n}(q_\mu\,;\,s,\,t) \equiv \{iq_3(1-st)+iq_1(s+t)+q_2(s-t)+q_4(1+st)\}^{2n} \tag{136}$$

by

$$G_{2n}(q_\mu\,;\,s,\,t) = \sum_{j,k=0}^{2n}\binom{2n}{j}H_{2n}^{(j,k)}(q_\mu)t^j s^k. \tag{137}$$

Explicit expressions for the $H_{2n}^{(r,l)}$ and a modified form of equation (135) (in polar coordinates) will be given in subsection 5.

Equation (135) can be deduced from (136) in the same way that the statement $\Delta h_{n,l}=0$ was deduced from (120). To prove equations (134) and (136) we use (124) and consider the effect of the linear transformation (131) upon the $h_{n,l}$. From equation (124) and equations (133) and (131) we find

$$h_{n,l}(x',\,y',\,z') = (-1)^n\binom{2n}{n+l}w_1'^{n+l}w_2'^{n-l}, \tag{138}$$

$$h_{n,l}(x',y',z', = (-1)^n\left(\frac{w_2}{v}\right)^{2n}\binom{2n}{n+l} \{-q_2+iq_1+(q_4-iq_3)s\}^{n+l}\{q_4+iq_3+(q_2+iq_1)s\}^{n-l} \tag{139}$$

where $s = w_1, w_2$. On the other hand, the right side in equation (136) is

$$\{[q_4+iq_3+(iq_1+q_2)s]+[iq_1-q_2+(q_4-iq_3)s]t\}^{2n};$$

therefore, the coefficient of $(-1)^n(w_2/v)^{2n}$ in (139) is precisely the coefficient of t^{n+l} in (137) with $s=w_1/w_2$. We obtain

$$h_{n,l}(x',y',z') = (-1)^n\binom{2n}{n+l}v^{-2n}\sum_{k=0}^{2n}H_{2n}^{(n+l,k)}w_1^k w_2^{2n-k}, \tag{140}$$

which together with (124) proves equation (134).

5. Expressions in Terms of Angular Coordinates

In polar coordinates the standard spherical harmonics express themselves in terms of r, θ, φ as products of functions of a single variable; however, the transformation formulas become simple in Cartesian coordinates. Indeed, if we set

$$x' = r'\cos\varphi'\sin\theta', \qquad y' = r'\sin\varphi'\sin\theta', \qquad z' = \cos\theta', \tag{141}$$

we have $r' = r$, but θ', φ' cannot be expressed in a simple manner in terms of θ and φ. Nevertheless, the spherical harmonics

$$(r')^{-n}h_{n,l}(x', y', z') = (-1)^n \frac{2^n n!}{(n+l)!} P_{n,l}(\cos\theta')e^{il\varphi'}$$

can be expressed in terms of the $P_{n,l}(\cos\theta)\exp\{il\varphi\}$ if we use (134). Only the problem of finding simple explicit expressions for the $H_{2n}^{(n+l,n+r)}(q_\mu)$ remains. If we were to introduce four-dimensional (hyperspherical) polar coordinates in the space of the variables q_μ, we would still obtain rather complicated formulas; however, there exists a particular coordinate system in which the expressions for the $H_{2n}^{(n+l,n+r)}$ become fairly simple. We may summarize the results as follows:

Let $v \geq 0$, $0 \leq \rho \leq 2\pi$, $0 \leq \sigma \leq 2\pi$, $0 \leq \tau \leq \pi/2$ *and*

$$\begin{aligned} q_1 &= -v\sin\sigma\sin\tau, & q_2 &= v\cos\sigma\sin\tau \\ q_3 &= v\sin\rho\cos\tau, & q_4 &= v\cos\rho\cos\tau. \end{aligned} \tag{142}$$

Then v, ρ, σ, τ *are orthogonal coordinates; the line element is*

$$dq_1^2 + dq_2^2 + dq_3^2 + dq_4^2 = dv^2 + v^2\{\cos^2\tau d\rho^2 + \sin^2\tau d\sigma^2 + d\tau^2\} \tag{143}$$

and the surface element of the hypersphere Ω *(defined by* $v = 1$*) is*

$$d\Omega = \cos\tau\sin\tau\, d\rho\, d\sigma\, d\tau. \tag{144}$$

In these coordinates

$$\begin{aligned} v^{-2n}H_{2n}^{(n+l,n+r)}(q_\mu) &= S_{2n}^{(l,r)}(\rho, \sigma, \tau) \\ &= \frac{1}{(n+r)!} e^{-i(l+r)\rho}e^{-i(r-l)\sigma}(\cos\tau)^{l+r}(\sin\tau)^{r-l}F_{2n}^{(l,r)}(\cos^2\tau) \end{aligned} \tag{145}$$

where

$$F_{2n}^{(l,r)}(t) = \frac{d^{n+r}}{dt^{n+r}} t^{n-l}(t-1)^{n+l}. \tag{146}$$

The $S_{2n}^{(l,r)}$ *and their conjugate complex functions are biorthogonal on the unit-hypersphere* Ω:

$$\iiint_\Omega S_{2n}^{(l',r')}\overline{S}_{2n}^{(l,r)}\, d\Omega = \begin{cases} 0 & \text{if } l' \neq l \text{ or } r' \neq r \\ \dfrac{2\pi^2}{2n+1}\binom{2n}{r+n}\Big/\binom{2n}{l+n} & \text{if } l' = l, \text{ and } r' = r. \end{cases} \tag{147}$$

The transformation formulas of the spherical harmonics are

$$P_{n,l}(\cos\theta')e^{il\varphi'} = \sum_{r=-n}^{n} \frac{(n-r)!}{(n-l)!} S_{2n}^{(l,r)} P_{n,r}(\cos\theta)e^{ir\varphi}. \tag{148}$$

Proof: If in (136) and (137) we substitute the expressions (142) for the q_ν and introduce s^* and t^* by

$$s^* = se^{-i(\sigma+\rho)}, \qquad t^* = te^{-i(\rho-\sigma)}$$

we find, for the $S_{2n}^{(l,r)}$ in (145),

$$\begin{aligned} &\{(1 + s^*t^*)\cos\tau + (s^* - t^*)\sin\tau\}^{2n} \\ &\qquad = \sum_{l,r=-n}^{n} \binom{2n}{n+l} S_{2n}^{(l,r)} e^{i(\tau+l)\rho} e^{i(\tau-l)\sigma} t^{*n+l} s^{*n+r} \end{aligned} \tag{149}$$

which proves that $S_{2n}^{(l,r)} \exp\{i[(r+l)\rho + (r-l)\sigma]\} = U_{2n}^{(l,r)}$ depends only on τ. Comparing the coefficients of t^{*n+l} on both sides of (149), we find

$$\begin{aligned} \sum_{r=-n}^{+n} U_{2n}^{(l,r)} s^{*n+r} &= (\cos\tau + s^*\sin\tau)^{n-l}(-\sin\tau + s^*\cos\tau)^{n+l} \\ &= (\cos\tau)^{l-n}(\sin\tau)^{-n-l}[\cos^2\tau + s^*\cos\tau\sin\tau]^{n-l} \\ &\qquad \cdot[\cos^2\tau - 1 + s^*\cos\tau\sin\tau]^{n+l}. \end{aligned} \tag{150}$$

Application of Taylor's formula to the last line in (150) yields equations (145) and (146). From these formulas and from (144) we can derive (147) by repeated integration by parts if we observe that

$$\bar{S}_{2n}^{(l,r)} = (-1)^{l+r} S_{2n}^{(-l,-r)}.$$

We mention without proof that, according to a formula which was discovered by Jacobi, the $U_{2n}^{(l,r)}$ can also be expressed in terms of hypergeometric series (see Chapter II, equation (24) on page 90). These series are, for $l + r \leq 0$,

$$\begin{aligned} U_{2n}^{(l,r)} &= (-1)^{n+l}\binom{n-l}{n+r}(\cos\tau)^{-l-r}(\sin\tau)^{l-r} \\ &\qquad \cdot {}_2F_1(-n-r, n+1-r; 1-l-r; \cos^2\tau), \end{aligned}$$

and, for $r + l \geq 0$,

$$\begin{aligned} U_{2n}^{(l,r)} &= (-1)^{n+r}\binom{n+l}{n-r}(\cos\tau)^{l+r}(\sin\tau)^{r-l} \\ &\qquad \cdot {}_2F_1(r-n, n+r+1; 1+l+r; \cos^2\tau). \end{aligned}$$

Additional Bibliography

CHAPTER I

Birkhoff, G., and MacLane, S. A survey of modern algebra. Macmillan, New York, 1941.

Frazer, R. A., Duncan, W. J., and Collar, A. R. Elementary matrices. Cambridge, England, the University Press, 1938.

Halmos, P. R. Finite dimensional vector spaces. Princeton, Princeton University Press, 1942 (Annals of mathematics studies no. 7).

Hamburger, H. L., and Gremshaw, M. E. Linear transformations in n-dimensional vector space. Cambridge, Eng., the University Press, 1951.

MacDuffee, C. C. Theory of matrices. J. Springer, Berlin, 1933 (Ergebnisse der Mathematik und ihrer Grenzgebiete, bd. 2, bd. 5).

Schreier, O., and Sperner, E. Einfuehrung in die analytische Geometrie und Algebra. B. G. Teubner, Leipzig und Berlin, 1935 (Hamburger mathematische Einzelschriften, Heft 10, Heft 19).

Turnbull, H. W., and Aitken, A. C. An introduction to the theory of canonical matrices. Blackie and Son, Ltd., London and Glasgow, 1932.

Wedderburn, J. H. M. Lectures on matrices. American Mathematical Society, New York, 1934 (Colloquium publications, vol. 17).

CHAPTER II

Bochner, S. Vorlesungen ueber Fouriersche Integrale. Akademische Verlagsgesellschaft, Leipzig, 1932.

Bochner, S., and Chandrasekharan, K. Fourier transforms. Princeton University Press, Princeton, 1949 (Annals of mathematics studies, no. 19).

Carleman, T. L'Intégrale de Fourier et questions qui s'y rattachent. Almquist and Wiksells boktr., Upsala, 1944.

Hardy, G. H., and Rogosinski, W. W. Fourier series. Cambridge, Eng., the University Press, 1950 (Cambridge tracts in mathematics and physics, no. 38).

Jackson, D. Fourier series and orthogonal polynomials. Oberlin, Ohio, 1941 (Carus mathematical monographs, no. 8).

Jeffreys, H., and Jeffreys, B. S. Methods of mathematical physics. Cambridge, Eng., the University Press, 1946.

Kaczmarz, S., and Steinhaus, H. Theorie der Orthogonalreihen. Z subwencji Funduszu kultury narodowej (Monografje matematyczne, vol. 6) Warsawa-Lwow, 1935.

Paley, R. E. A. C., and Wiener, N. Fourier transforms in the complex domain. American Mathematical Society, New York, 1934 (Colloquium publications, vol. 19).

Rogosinski, W. Fouriersche Reihen. W. de Gruyter, Berlin, 1930.

Sneddon, I. N. Fourier transforms. McGraw-Hill, New York, 1951.

Szegö, G. Orthogonal polynomials. American Mathematical Society, New York, 1939 (Colloquium publications, vol. 23).
Titchmarsh, E. C. Introduction to the theory of Fourier integrals. The Clarendon Press, Oxford, 1937.
Wiener, N. The Fourier integral and certain of its applications. Cambridge, Eng., the University Press, 1933.
Zygmund, A. Trigonometrical series. Z subwencji Funduszu kultury narodowej (Monografje matematyczne, vol. 5), Warsawa-Lwow, 1935.

CHAPTER III

Banach, S. Théorie des opérations linéaires. Z subwencji Funduszu kultury narodowej (Monografje matematyczne, vol. 1), Warsawa, 1932.
Hamel, G. Integralgleichungen. J. Springer, Berlin, 1937.
Lovitt, S. V. Linear integral equations. McGraw-Hill, New York, 1924.
Riesz, F. Les systèmes d'équations linéaires à une infinité d'inconnues. Gauthier-Villars, Paris, 1913.
Stone, M. H. Linear transformations in Hilbert space and their application to analysis. American Mathematical Society, New York, 1932 (Colloquium publications, vol. 15).

CHAPTER IV

Bliss, G. A. Lectures on the calculus of variations. University of Chicago Press, Chicago, 1946.
Carathéodory, C. Variationsrechnung und partielle Differentialgleichungen erster Ordnung. B. G. Teubner, Leipzig and Berlin, 1935.
Courant, R. Dirichlet's principle, conformal mapping, and minimal surfaces. Interscience, New York, 1950.
Forsyth, A. R. Calculus of variations. Cambridge, Eng., the University Press, 1927.
Fox, C. An introduction to the calculus of variations. Oxford University Press, London, 1950.
Lanczos, C. The variational principles of mechanics. University of Toronto Press, Toronto, 1949.
Lusternik, L., and Schnirelmann, L. Méthodes topologiques dans les problèmes variationnels. Hermann et Cie., Paris, 1934.
Lusternik, L., and Schnirelmann, L. Functional topology and abstract variational theory. Paris, Hermann and Cie., 1938.
Morrey, C. B. Multiple integral problems in the calculus of variations and related topics. University of California Press, Berkeley and Los Angeles, 1943.
Morse, M. The calculus of variations in the large. American Mathematical Society, New York, 1934 (A.M.S. Colloquium publications v. 18).
Seifert, H., and Threlfall, W. Variationsrechnung im Grossen. B. G. Teubner, Leipzig and Berlin, 1938. (Hamburger mathematische Einzelschriften Heft 24)

Polya, G., and Szegö, G. Isoperimetric inequalities in mathematical physics. Princeton University Press, 1951 (Annals of mathematics studies, 27).

CHAPTER IV (Papers)

Courant, R. Variational methods for the solution of problems of equilibrium and vibrations. American Mathematical Society, Bulletin, vol. 49, pp. 1–23, 1943.

CHAPTERS V AND VI

Churchill, R. V. Fourier series and boundary value problems. McGraw-Hill, New York, 1941.

Collatz, L. Eigenwertprobleme und ihre numerische Behandlung. Becker und Erler. Leipzig, 1945.

Hohenemser, K., Die Methoden zur angenäherten Lösung von Eigenwertproblemen in der Elastokinetik. Chelsea, New York, 1949.

Ince, E. L., Ordinary differential equations. Dover, New York, 1944.

Rellich, F., Störungstheorie der Spektralzerlegung. J. Springer, Berlin, 1936–42.

Titchmarsh, E. C. Eigenfunction expansions associated with second-order differential equations. Oxford, the Clarendon Press, 1946.

Weyl, H. Ramifications, old and new, of the eigenvalue problem. Bulletin of the American Mathematical Society, Volume 56, 1950, pp. 115–139.

Weinstein, A. Étude des spectres des équations aux derivées partielles de la théorie des plaques élastiques. Mémorial des Sciences Mathématiques, Volume 88, 1937.

Proceedings of the Symposium on Spectral Theory and Differential Problems. Oklahoma A. & M., Stillwater, Oklahoma, 1951.

CHAPTER VII

Gray, A., Matthews, G. B., and MacRobert, T. M. A treatise on Bessel functions and their applications to physics. Macmillan, London, 1922.

Hobson, E. W. The theory of spherical and ellipsoidal harmonics. Cambridge, England, the University Press, 1931.

Jahnke, E., and Emde, F. Tables of functions with formulae and curves. Dover, New York, 1943.

MacRobert, T. M. Spherical harmonics: an elementary treatise on harmonic functions. London, Methuen and Co., 1927.

Madelung, E. Die mathematischen Hilfsmittel des Physikers. Berlin, J. Springer, 1936 (Die Grundlehren der mathematischen Wissenschaften in Einzeldarstellungen, Bd. 4).

Magnus, W., and Oberhettinger, F. Formeln und Sätze für die speziellen Funktionen der mathematischen Physik. Berlin, J. Springer, 1948 (Die Grundlehren der mathematischen Wissenschaften in Einzeldarstellungen, Bd. 52).

Mclachlan, N. W. Bessel functions for engineers. Oxford, the Clarendon Press, 1934.

Relton, F. E. Applied Bessel functions. London, Blackie and Son Ltd., 1946.
Watson, G. N. A treatise on the theory of Bessel Functions. Cambridge, England, the University Press, 1944; New York, Macmillan, 1944.
Weyrich, R. Die Zylinderfunktionen und ihre Anwendungen. Leipzig und Berlin, B. G. Teubner, 1937.

Papers

Copson, E. T. The asymptotic expansion of a Function defined by a definite integral or contour integral, Admiralty Computing Service.
Van der Waerden, B. L. On The method of saddle points, Applied Scientific Research B (The Hague), vol. II, No. 1, 1951, pp. 33–45.

INDEX

A

Abel's integral equation, 158
Accumulation principle for functions, 57
Adjoint differential expression, 279, 280
Adjoint orthogonal systems of unsymmetric kernels, 159
Alternative theorem for integral equations, 115
  for linear equations, 6
  for ordinary differential equations, 359
Analytic continuation, 395
Approximation in the mean, 52, 54
Approximation theorem of Weierstrass, 65
Argument function, 168
Arzela's theorem, 59
Associated Legendre functions, 327, 505
Asymptotic behavior
  of eigenvalues, in Sturm-Liouville problem, 414
    of Bessel's equation, 415
    of vibrating plate, 460
  of solutions of the Sturm-Liouville problem, 331–339
Asymptotic dimension of a sequence of functions, 63
Asymptotic distribution of eigenvalues of a differential equation, 429–445, 460
Asymptotic expansions, 522–535
  for Bessel and Hankel functions, 524, 532

B

Beats, 390
Bessel's equation, 303, 467
Bessel's equation (*continued*)
  asymptotic behavior of eigenvalues of, 415
  singularities of solutions of, 501
Bessel functions 303 306, 324 325, 340, 372, 390–391, 402–403, 467 501
  asymptotic behavior of, 332 4
  asymptotic representations of, 524 532
  functional relation for, 489
  integral representations for, 474–82
  integral theorem for, 340–341, 490
  orthogonality relations for, 305
  power series for, 482 484
  relations between, 485
  singularities of, 501
  zeros of, 451, 492–496
Bessel's inequality
  for functions, 51, 425
  for vectors, 5
Bessel transformation, 490–492
Beta function, 483
Bezout's theorem, 520
Biharmonic equation, 370
Bilinear forms, 11 14
  reciprocal, 12
  resolvents of, 18, 29
Bilinear formula, for iterated kernels, 137
Bilinear integral form, 123
Bilinear relation, 360
Biorthogonality relations, 404
du Bois-Reymond, theorem of, 200
Boundary, free, 208
Boundary conditions
  containing a parameter, 392–393, 461
  homogeneous and inhomogeneous, 277
  natural, 208

Boundary conditions (*continued*)
for the rod, 295
for the string, 291
Brachistochrone, 170, 207
Broken extremals, 259
Buckling, variational problem of, 272

C

Calculus of perturbations, 344
Calculus of variations, 164–274
applied to eigenvalue problems 397–465
fundamental lemma of, 185
Canonical differential equations, 239
Canonical form of variational problems, 238
Castigliano's principle, 270
Catenary, 172, 218
Cayley's theorem, 22, 536–539
Characteristic values of a matrix, 17
Closed systems of functions, 110
Complete systems of functions, 52, 54
Completeness
of eigenfunctions of a differential equation, 359–360, 369, 424
of Hermite polynomials, 96–97
of Laguerre polynomials, 96
of Legendre polynomials, 82
of powers of x, 65, 101–102
of spherical harmonics, 512 513
of a system of functions, 110
of systems of functions of several variables, 56
of systems of vectors, 4
of trigonometric functions, 68
Completeness relation
for vectors, 5
for functions, 51
Components of a function, 51
Conduction of heat, 311
Confocal rectangular parallelepiped, 319
Conformal mapping, 377
Conjugate potential functions, 241
Conservation of momentum, 265
Constrained systems, eigenvalues of, 408
Constraints
on quadratic forms, 33, 44
on variational problems, 233
Continued fraction for the cotangent, 489
Continuous dependence of eigenvalues and eigenfunctions on the kernel, 151
Continuous spectrum, 100, 339 343
Convergence in the mean, 98, 110
Convergence theorems of Lebesgue, 109–111
Cotangent, continued fraction for, 489
Cylinder functions, *see* Bessel, Hankel, Mathieu, Neumann functions
Cylindrical domain, 319

D

Darboux method, 532–535
Definite kernels, 123, 162
Degeneracy of eigenvalues, 129
Degenerate kernels, 114–118, 129
Degenerate quadratic forms, 27
Dense function systems, 100–101
Derivative, variational, 186
Determinant inequality of Hadamard, 36
Dido's problem, 258
Differential equations
canonical, 239
equivalence with integral equations, 358–363
integral transformations applied to, 466–468
Dimension number, asymptotic, 63
Dini's theorem, 57
Dipole, 514–15
Direct solution of variational problems, 174–183
Dirichlet's discontinuous factor, 81
Dirichlet's integral formula, 78
Dirichlet's problem, 240
for the circle, 179
Discontinuous factor of Dirichlet, 81
Discontinuous kernels, 152

Discrete spectrum, 130
Divergence expressions, 195
Du Bois-Reymond, theorem of, 200

E

Eigenfrequency, 283
Eigenfunctions
  continuous dependence on kernel, 151
  of a differential equation, nodes of, 451–455
    completeness of, 424
  Kellogg's method for, 156
  of a kernel, 113
    minimum properties of, 161
  nodes of, 300, 302, 304, 395, 451–455, 458, 463
  for rectangular membrane, 301
  of symmetric kernels, 126–132, 148–150
  zeros of, 451–455
Eigenvalues
  of Bessel's equation, asymptotic behavior of, 415
  of a constrained system, 408
  continuous dependence on kernel, 151
  degenerate, 129
  of a differential equation, asymptotic distribution of, 429–445, 460
    continuity of, 418–424
    dependence on boundary conditions, 410–411
    dependence on coefficients, 411
    extremum properties of, 398–407
    infinite growth of, 412
    maximum-minimum property of, 405
    minimum properties of, 459
    for a subdomain, 408–409
  of functions of a kernel, 157
  of functions of a matrix, 22
  of infinite multiplicity, 154, 395
  of a kernel, 113, 132, 137
  of a matrix, 17, 26
  for the membrane, 298
  minimum-maximum property, 31, 132–134, 405
  multiple, 113, 129
  negative, 294, 416
  for the plate, asymptotic behavior of, 460
  of a positive definite kernel, 132
  reciprocal, 17
  for the rod, 296
  for the string, 287
  of Sturm-Liouville problem, asymptotic behavior of, 414
  of a symmetric kernel, 148–150
    first, 122–125
    maximum-minimum property of, 132–134
    totality of, 126–132
  of a unitary matrix, 46
Eigenvalue problems
  for closed surfaces, 461
  definition of, 309
  of mathematical physics, 275–396
  with two parameters, 460
  of Schrödinger type, 445–450
  variational methods for, 397–465
Eigenvectors, 24
  use in solving linear equations, 30
Eigenvibrations, 283, 310
  of homogeneous string, 287
Elastic bodies, 268
Electrostatics, 267
Elementary divisors of a matrix, 45
Ellipsoidal coordinates, 226
Elliptic coordinates, 319, 391–392
Elliptic functions, 227
Energy
  kinetic, 242
  potential, 242
Energy integral, 265
Enskog's method for symmetric integral equations, 156
Equicontinuous functions, 58, 114, 119, 125
Equilibrium conditions of a mechanical system, 243
Equilibrium problems, 308
Erdmann vertex condition, 260

Euler equations, 183–207
homogeneous form of, 196
integration of, 206
invariance of, 222
for isoperimetric problem, 207
Euler expression, identical vanishing of, 193
Euler integral, 483
Euler transformation, 467
Expansions, asymptotic, 522–535
Expansion theorem, 310
for definite kernels, 138–140
for differential equations, 360–361, 427–428
for iterated kernels, 137
limits of validity of, 395–396
for spherical harmonics, 513
for symmetric integral equations, 134–136
Extrema, Weierstrass's theorem on, 20, 164
Extremals, 185
broken, 259
Extremum properties of eigenvalues, 398–407

F

Fejér's summation theorem, 102
Fermat's principle, 165
Finite differences, 176
First variation, 184, 208–14
Fischer-Riesz theorem, 110
Forced motion
of the membrane, 300
of the string, 289, 294
Forced vibrations, 285, 389–390
Forms, *see under* particular type of form; e.g., bilinear, quadratic, Hermitian, etc.
Form factor, 286
Fourier coefficients, 51, 424
order of magnitude of, 74
Fourier integral, 77–82, 98–100
Fourier integral theorem, 78
Fourier series, 69–77
Fredholm formulas, 142–147
Fredholm's theorems
for arbitrary kernels, 118–122, 150
for degenerate kernels, 115–118
Schmidt's derivation of, 155
Free boundaries, 208
Free vibrations, 282–283
Friedrichs transformation, 234
Function space, 57
gradient in, 222
Functions of a kernel, eigenvalues of, 157
Functions of a matrix, eigenvalues of, 22
Functionals, 167–169
Fundamental domain, 48, 112
Fundamental lemma of the calculus of variations, 185
Fundamental solution, 370
of an ordinary differential equation, 353
Fundamental tone, 286

G

Generalized Green's function, 356–357
Generating function
for Hermite and Laguerre polynomials, 510
for spherical harmonics, 539–541
Geodesic curves, 170, 189, 213
Gibbs' phenomenon, 105–107
Gradient in function space, 222
Gram determinant
of a system of functions, 62, 107–108
of a system of vectors, 34–36
Gram-Schmidt orthogonalization process, 4, 50
Green's formula, 278, 280
Green's function, 311, 351–388
for the biharmonic equation, 370
for a circular ring, 386–388
obtained by conformal mapping, 377
construction by reflection, 378–379, 384–385
examples of, 371–388
in the generalized sense, 356–357

for Hermite's equation, 373–375
as kernel of an integral equation, 358
for Laguerre's equation, 375–376
for ordinary differential equations, 351–363
construction of, 354–355
definition of, 352–353
of higher order, 362–363
for the potential equation, 363–370, 377
in the plane, 384–386
in space, 378–384
symmetry of, 354
Green's tensors for systems of differential equations, 393–394
Group, orthogonal, 539

H

Haar, theorem of, 203
Hadamard's inequality for determinants, 36
Hamilton's variational principle, 242
Hammerstein's theorem, 162
Hankel functions, 468–473
asymptotic representations for, 524–532
integral representation of, 476
singularities of, 501
Harmonic functions, *see* potential equation
Heat conduction, equation of, 311
Herglotz, theorem of, 542
Hermite equation, 508
Green's function for, 373–375
Hermite functions, 508–509
Hermite orthogonal functions, 154
Hermite polynomials, 91–93, 328
completeness of, 96–97
generating function for, 510
Hermitian forms, 13, 28
Hilbert space, 55
orthogonal transformations in, 55
Holonomic conditions, 221
Hurwitz's solution of the isoperimetric problem, 97–98

I

Independence, measure of, 61
Indicatrix, 258
Inertial theorem for quadratic forms, 28
Infinitely many variables, 168, 177
applied to integral equations, 160
Infinitesimal linear transformations, 41–42
Influence function, *see* Green's function
Inner product
of functions, 49
of vectors, 1
Integral equation of Abel, 158
Integral equations (linear), 112–163
alternative theorem for, 115
with discontinuous kernels, 152
Enskog's method for, 156
equivalence with differential equations, 358–363
Fredholm theorems for, 115–122, 150
for functions of several variables, 152
homogeneous, 112
inhomogeneous, solution of, 136
of the first kind, 159
of the second kind, 112
of the third kind, 161
polar, 161
reciprocal, 141
singular, 154
symmetric, 156
transposed, 116
Volterra, 158, 337–339
Integral formula of Dirichlet, 78
Integral representations
of Bessel functions, 474–482
of Hankel functions, 476
of Hermite function, 508–509
of Laguerre functions, 509–510
of Legendre functions, 502–505
of Neumann functions, 498
of Tchebycheff functions, 507–508

Integral theorem
for Bessel functions, 340–341, 490
of Fourier, 78
Integral transforms, 113
Integral transformations
applied to differential equations, 466–468
definition of, 112
of Mellin, 104
Integro-differential equations, 404
Invariance of Euler equations, 222
Invariant variational problems, 262
Involutory transformation of variational problems, 233
Isoperimetric problem(s), 171–172, 216
Euler equation of, 207
Hurwitz solution of, 97
for plane curves, 174
for polygons, 166
reciprocity for, 258
Iterated bilinear form, 12
Iterated kernels, 137

J

Jacobi's method of multiplicative variation, 458
Jacobi polynomials, 90–91, 327–328
Jump conditions, 404

K

Kellogg's method for eigenfunctions, 156
Kernels
definite, 123, 162
definition of, 112
degenerate, 114–118, 129
discontinuous, 152
eigenfunctions of, 113
eigenvalues of, 113, 132, 137
expansion theorem for, 134–140
functions of, 157
iterated, 137
positive definite, 132
possessing no eigenvalues, 157, 158
reciprocal, 141
resolvents of, 141–147, 162
symbolic functions of, 157
symmetric, *see* Symmetric kernels
symmetrizable, 161
unsymmetric, adjoint orthogonal systems of, 159
Kinetic energy, 242

L

Lagrange's equations of motion, 243
Lagrange multipliers, 165
Laguerre's equation, Green's function for, 375–376
Laguerre functions, 509–510
Laguerre orthogonal functions, 154
Laguerre polynomials, 93–97, 329–330, 342–343
completeness of, 95
generating function for, 510
Lamé equation, 320
Lamé functions, 320
Lamé problem, 319–324
Laplace's equation, *see* potential equation
Laplace integral, 82
Laplace's integral representation for Legendre functions, 503–504
Laplace spherical harmonics, 316, 510–521
Laplace transformation, 467
Laplacian, transformation to polar coordinates, 224
Least time, principle of, 165
Lebesgue convergence theorem, 109–111
Lebesgue integral, 108–111
Legendre condition, 186, 200, 202, 214
strong form of, 216
Legendre functions, 325–327, 501–507
associated, 327, 505
integral representations for, 502–505
of the second kind, 504–505
Legendre polynomials, 82–87, 402
asymptotic expansion of, 533–535

completeness of, 82
differential equation for, 86
generating function for, 85
of higher order, 325, 505
minimum property of, 86
recursion formula for, 86
as special spherical harmonics, 318
Light rays, 165, 170, 189, 213, 258
Linear dependence
of functions, 50
of vectors, 2
Linear differential operator, 275
Linear equations, 5–6
perturbations of, 42–44
solution by means of eigenvectors, 30
Linear transformations, 1–47
infinitesimal, 41–42

M

Mathematical physics, problems of, 275–396
Mathieu functions, 392
Matrices, 7–11
characteristic values of, 17
eigenvalues of, 17, 26
elementary divisors of, 45
functions of, 22
spectra of, 17
Maximizing sequences, 175
Maximum-minimum property
of eigenvalues of a differential equation, 405
of nodal points, 463
Maxwell-Sylvester representation for spherical harmonics, 514–521
Mean, approximation in the, 52, 54
Mean convergence, 98, 110
Mean square error, 52
Measure of independence, 61
Measure of a set, 108
Mellin transformation, 103–105
Membrane, 246, 249, 297–307, 463
circular, 302–306
eigenvalues for, 298
forced motion of, 300
non-homogeneous, 306
rectangular, 300–302
Mercer's theorem, 138–140
application to differential equations, 360, 363, 369
Minimal surfaces, 193, 258
Minimizing sequences, 175
Minimum property
of eigenfunctions, 161
of eigenvalues, derived from completeness, 456
of Legendre polynomials, 86
Minimum-maximum principle, for quadratic forms, 31
for symmetric kernels, 132–134
Momentum, conservation of, 265
Multiple Fourier series, 73
Multiplicative variation, Jacobi's method of, 458
Multiplicity of an eigenvalue, 113, 129
Multipliers, Lagrange, 165
Multiplier rule for variational problems, 222
Multipole, 515
Müntz's theorem, 101–102

N

Natural boundary conditions, 208
Neighborhood of a function, 169
Neumann functions, 471–473, 496–501
singularities of, 501
Neumann series, of bilinear forms, 18
for integral equations, 140–142
of matrices, 10
Nodal lines, 300, 302, 304, 395
Nodal points for Sturm-Liouville problem, 454, 463
Nodes, 451–454, 458
Noether's theorem, 262
Norm
of a function, 49
of a vector, 2
Normal coordinates, 282
Normal modes, 283
Null sets, 108

O

Orthogonal function system, of Hermite, 154
of Laguerre, 154
Orthogonal group, 539
Orthogonal systems
of functions, 49–50, 87
of vectors, 3–5
Orthogonal transformations, 14–16, 536
in Hilbert space, 55
Orthogonalization
of functions, 50
of vectors, 4
Orthonormal system
of functions, 49
of vectors, 3
Oscillation theorem, 454
Overtones, 286

P

Paraboloidal coordinates, 229
Parallelepiped, confocal, 319
Parameters in boundary conditions, 392–393, 461
Particle mechanics, integrals in, 265
Perturbations, 42
Perturbation theory, 343–350
Picard's theorem, 160
Piecewise continuous functions, 48
Piecewise smooth functions, 48
Pitch, 283
Plate, 250–251, 307–308, 403, 463
asymptotic behavior of eigenvalues for, 460
Plucked string, 388
Point force, 351
Poisson's equation, 368
Poisson's integral, 513–14
Poisson's summation formula, 76
Polar form of a quadratic form, 12
Polar integral equations, 161
Polynomials, completeness of, 65, 101–102
Potential energy, 242
Potential equation
Green's function for, 363–370, 377
in the plane, Green's function for, 384–6
in space, Green's function for, 378–384
transformation to polar coordinates, 224
Potential functions, conjugate, 241
Potential theory, 178–182, 315–324, 363–371, 377–388
Principal axes transformations, of quadratic forms, 23–31
of two quadratic forms simultaneously, 37–41
Principle of Castigliano, 270
Principle of Fermat, 165
Propagation of light, 170, 189, 214
Pythagorean theorem for function spaces, 253

Q

Quadratic forms, 12–16, 23–34
constraints on, 33, 44
degenerate, 27
inertial theorem for, 28
polar forms of, 12
simultaneous transformation to principal axes, 37–41
transformation to principal axes, 23–31
Quadratic integral form, 122
Quaternions, 539

R

Rayleigh-Ritz method, 175
Reciprocal bilinear form, 12
Neumann series for, 19
Reciprocal eigenvalue, 17
Reciprocal kernel, 141
Reciprocal spectrum, 17
Reciprocal variational problems, 235, 252–257, 270
Reciprocity for isoperimetric problems, 258
Reciprocity formulas for Fourier integrals, 80

Rectangular parallelepiped, confocal, 319
Recursion formula for Legendre polynomials, 86
Reflection, law of, 165
Reflection method of constructing Green's function, 378–379, 384–385
Refraction, law of, 166
Regions with separate components, 405
Rellich, lemma of, 414, 450
Resolvent
  of a bilinear form, 18, 29
  of a kernel, 141–147, 162
  of a symmetric kernel, 142
Resonance, 285
Riesz-Fischer theorem, 110
Ritz, *see* Rayleigh-Ritz
Rod, 245, 295–297
  boundary conditions for, 295
  eigenvalues for, 296

S

Saddle point method, 526–533
Schläfli's integral, 501–502
Schmidt method for derivation of Fredholm theorem, 155
Schmidt orthogonalization process, *see* Gram-Schmidt
Schwarz inequality
  for functions, 49
  for vectors, 2
Schrödinger eigenvalue problems, 445–450
Schrödinger's equation, 342–343
Second variation, 214
Self-adjoint differential expression, 279
Separated regions, 405
Separation of variables, 287, 313
Series
  Fourier, 69–77
  Neumann, 10, 18, 140–142
Series expansions of arbitrary functions, 48–111
Singular equations, eigenvalues of, 461
Singular integral equations, 154
Small vibrations, 244
Smooth sequences of functions, 63
Snell's law of refraction, 166
Spectral decomposition, 99–100
Spectral density, 100
Spectrum, 339–340
  continuous, 100, 339–343
  discrete, 130
  of a matrix, 17
  reciprocal, 17
  of a unitary matrix, 46
Spherical coordinates, 225
Spherical harmonics, 314, 316, 342–343, 510–521
  completeness of, 512–513
  expansion theorem for, 513
  generating function for, 539–541
  Maxwell-Sylvester representation for, 514–521
  symmetric, 512
  transformations of, 535–545
Spheroidal coordinates, 229
Stationary functions and curves, 186
Stationary values in variational problems, 186
Steepest descent, method of, 526–533
Steiner's problem, 166
Stirling's formula, 522–524
String, 244, 388–392, 463
  boundary conditions for, 291
  eigenvalues for, 287
  free motion of, 287
  forced vibrations of, 289, 389
  impulsive excitation of, 389
  non-homogeneous, 291
  plucked, 388
Sturm-Liouville problem, 291–295
  asymptotic behavior of eigenvalues of, 414
  asymptotic behavior of solutions of, 331–339
  eigenvalue problem for, 402
  singular case, 324
  zeros of eigenfunctions of, 468

Subsidiary conditions for variational problems, 216
Summation formula of Poisson, 76
Superposition, principle of, 276
Sylvester, 518
Symmetric integral equations, Enskog's method for, 156
Symmetric kernels
  definition of, 115
  eigenfunctions of, 126–132, 148–150
  eigenvalues of, 122–134, 148–150
  expansion theorem for, 134–136
  resolvents of, 142
Symmetrizable kernels, 161
Synchronism, 309
Synchronous vibrations, 286

T

Tchebycheff functions, 507–508
Tchebycheff polynomials, 88–90, 328
Tensors, 7
  Green's, 393–394
Thales, theorem of, 255
Theta function
  applications of, 381–384, 388
  functional equation for, 75
Thomson's principle, 267
Three-dimensional continua, vibrations of, 313
Tone, fundamental, 286
Transformation(s)
  Bessel, 490–492
  of Euler, 467
  integral, 112
  linear, 1–47
    infinitesimal, 41–42
  orthogonal, 14–16
  of quadratic forms to principal axes, 23–31, 37–41
  unitary, 16
  of variational problems, 231–242
Transforms, integral, 113
Transposed integral equation, 116
Transversality, 211
  for multiple integrals, 266

U

Unitary matrix, 11
  spectrum of a, 46
Unitary transformations, 16
Unsymmetric kernels, adjoint orthogonal systems for, 159

V

Variation
  first, 186
  of a function, 184
  second, 214
Variational derivative, 186
Variational principle of Hamilton, 242
Variational problems
  canonical form of, 238
  constraints on, 233
  direct solution of, 174–183
  invariant, 262
  involutory transformation of, 233
  multiplier rule for, 222
  reciprocal, 235, 252–257, 270
  with subsidiary conditions, 216
  transformations of, 231–42
  with variable domains, 260
  without a solution, 173
Variations, calculus of, 164–274
Vectors, 1–5
Vertex condition of Weierstrass and Erdmann, 260
Vibrating membrane, *see* Membrane
Vibrating plate, *see* Plate
Vibrating rod, *see* Rod
Vibrating string, *see* String
Vibration problems of mathematical physics, 275–396
Vibrations
  forced, 285, 389–390
  free, 282–283
  small, 244
Volterra integral equations, 158, 337–339

W

Wave front, 214
Weierstrass-Erdmann vertex condition, 260
Weierstrass's approximation theorem, 65
Weierstrass's theorem on extrema, 20, 164
Weight function, 87

Z

Zeros of Bessel functions, 451, 492–496
Zeros of eigenfunctions, 451–455